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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission.

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


Founding Editor
Kinshuk, Athabasca University, Canada

Editors
Nian-Shing Chen, National Sun Yat-sen University, Taiwan; Demetrios G Sampson, Curtin University, Australia.

Advisory Board
Ignacio Aedo, Universidade Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohammed Jenni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanassis Karoulis, Hellenic Open University, Greece; Paul Kirschs, 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; Rudy Le louche, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McGreal, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Vaxjo University, Sweden; Riiichiro 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 Tso, 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; Ashok Patel, CAL Research & Software Engineering Centre, UK; Reinhard Oppermann, Fraunhofer Institut Angewandte Informationstechnik, Germany; Vladimir A Fomichev, K. E. Tsiolkovsky 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.

Editorial Assistant
Sie Wai (Sylvia) Chew, National Sun Yat-sen University, Taiwan; Stelios Sergis, University of Piraeus & Centre for Research and Technology Hellas, Greece; Wei-Chieh Fang, National Sun Yat-sen University, Taiwan.

Technical Manager
Stelios Sergis, University of Piraeus & Centre for Research and Technology Hellas, Greece

Executive Peer-Reviewers
http://www.ifets.info/

Publisher
National Sun Yat-sen University, Taiwan

Editorial Office
c/o Dr Nian-Shing Chen, Department of Information Management, National Sun Yat-sen University, 70, Lien-Hai Rd, Kaohsiung, 80424, Taiwan.

ISSN 1436-4522 (online) and 1176-3647 (print). This article of the Journal of Educational Technology & Society is available under Creative Commons CC-BY-ND-NC 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Journal Editors at ets-editors@ifets.info.
# Table of Contents

## Full Length Articles

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designing and Implementing Web-based Scaffolding Tools for Technology-enhanced Socioscientific Inquiry</td>
<td>1–12</td>
</tr>
<tr>
<td><em>Suhkyung Shin, Thomas A. Brush and Krista D. Glazewski</em></td>
<td></td>
</tr>
<tr>
<td>Are They Thinking Differently: A Cross-Cultural Study on the Relationship of Thinking Styles and Emerging Roles in Computer-Supported Collaborative Learning</td>
<td>13–24</td>
</tr>
<tr>
<td><em>Xiaoqing Gu, Huawen Wang and Jon Mason</em></td>
<td></td>
</tr>
<tr>
<td>Philosophy of Technology Assumptions in Educational Technology Leadership</td>
<td>25–36</td>
</tr>
<tr>
<td><em>Mark David Webster</em></td>
<td></td>
</tr>
<tr>
<td>Mobile Learning in Pre-Kindergarten: Using Student Feedback to Inform Practice</td>
<td>37–44</td>
</tr>
<tr>
<td><em>Jennifer L. Reeves, Glenda A. Gunter and Candace Lacey</em></td>
<td></td>
</tr>
<tr>
<td>Exploring Learners’ Sequential Behavioral Patterns, Flow Experience, and Learning Performance in an Anti-Phishing Educational Game</td>
<td>45–60</td>
</tr>
<tr>
<td><em>Jerry Chih-Yuan Sun, Cian-Yu Kuo, Huei-Tse Hou and Yu-Yan Lin</em></td>
<td></td>
</tr>
<tr>
<td>Computerized Dynamic Adaptive Tests with Immediately Individualized Feedback for Primary School Mathematics Learning</td>
<td>61–72</td>
</tr>
<tr>
<td><em>Huey-Min Wu, Bor-Chen Kuo and Su-Chen Wang</em></td>
<td></td>
</tr>
<tr>
<td>Comment Data Mining to Estimate Student Performance Considering Consecutive Lessons</td>
<td>73–86</td>
</tr>
<tr>
<td><em>Shaymaa E. Sorour, Kazumasa Goda and Tsunenori Mine</em></td>
<td></td>
</tr>
<tr>
<td>Effects of the Team Competition-Based Ubiquitous Gaming Approach on Students’ Interactive Patterns, Collective Efficacy and Awareness of Collaboration and Communication</td>
<td>87–98</td>
</tr>
<tr>
<td><em>Chih-Hung Chen and Gwo-Jen Hwang</em></td>
<td></td>
</tr>
<tr>
<td>Conversational Agents Improve Peer Learning through Building on Prior Knowledge</td>
<td>99–111</td>
</tr>
<tr>
<td><em>Stergios Tegos and Stavros Demetriadis</em></td>
<td></td>
</tr>
<tr>
<td>Journalogue: Voicing Student Challenges in Writing Through a Classroom Blog</td>
<td>112–122</td>
</tr>
<tr>
<td><em>Suneeta Thomas</em></td>
<td></td>
</tr>
<tr>
<td>The Effect of Computer Game-Based Learning on FL Vocabulary Transferability</td>
<td>123–133</td>
</tr>
<tr>
<td><em>Stephan J. Franciosi</em></td>
<td></td>
</tr>
<tr>
<td>Surveying In-Service Teachers’ Beliefs about Game-Based Learning and Perceptions of Technological Pedagogical and Content Knowledge of Games</td>
<td>134–143</td>
</tr>
<tr>
<td><em>Chung-Yuan Hsu, Meng-Jung Tsai, Yu-Hsuan Chang and Jyh-Chong Liang</em></td>
<td></td>
</tr>
<tr>
<td>Wikis for a Collaborative Problem-Solving (CPS) Module for Secondary School Science</td>
<td>144–155</td>
</tr>
<tr>
<td><em>Dorothy DeWitt, Norlida Alias, Saedah Siraj and Jonathan Michael Spector</em></td>
<td></td>
</tr>
<tr>
<td>How Augmented Reality Enables Conceptual Understanding of Challenging Science Content</td>
<td>156–168</td>
</tr>
<tr>
<td><em>Susan Yoon, Emma Anderson, Joyce Lin and Karen Elinich</em></td>
<td></td>
</tr>
<tr>
<td>Online Research Behaviors of Engineering Graduate Students in Taiwan</td>
<td>169–179</td>
</tr>
<tr>
<td><em>Ying-Hsueh Cheng and Chin-Chung Tsai</em></td>
<td></td>
</tr>
</tbody>
</table>
Guest Editorial

“HOW” to Design, Implement and Evaluate the Flipped Classroom?
Yanjie Song, Morris Jong, Maiga Chang and Weiqin Chen

Special Issue Articles

Facilitating and Bridging Out-Of-Class and In-Class Learning: An Interactive E-Book-Based Flipped Learning Approach for Math Courses
Gwo-Jen Hwang and Chiu-Lin Lai

184–197

An Experiential Learning Perspective on Students’ Satisfaction Model in a Flipped Classroom Context
Xuesong Zhai, Jiabao Gu, Heji Liu, Jyh-Chong Liang and Chin-Chang Tsai

198–210

Implementing the Flipped Classroom in Teacher Education: Evidence from Turkey
Gökçe Kurt

211–221

Using “First Principles of Instruction” to Design Secondary School Mathematics Flipped Classroom: The Findings of Two Exploratory Studies
Chung Kwan Lo and Khee Foon Hew

222–236

An Action Research Study from Implementing the Flipped Classroom Model in Primary School History Teaching and Learning
Vasiliki Aidinopoulou and Demetrios G. Sampson

237–247

Conceptualizing “Homework” in Flipped Mathematics Classes
Zandra de Araujo, Samuel Otten and Salih Birisci

248–260

Investigating the Potential of the Flipped Classroom Model in K-12 ICT Teaching and Learning: An Action Research Study
Christoforos Kostaris, Stylianos Sergis, Demetrios G. Sampson, Michail N. Giannakos and Lina Pelliccione

261–273

Integrating the SOP2 Model into the Flipped Classroom to Foster Cognitive Presence and Learning Achievements
Hsiu-Ling Chen and Chiung-Yun Chang

274–291

How to Flip the Classroom – “Productive Failure” or “Traditional Flipped Classroom” Pedagogical Design?
Yanjie Song and Manu Kapur

292–305

Empowering Students in the Process of Social Inquiry Learning through Flipping the Classroom
Morris Siu-Yung Jong

306–322

Flipping Business Education: Transformative Use of Team-Based Learning in Human Resource Management Classrooms
Chung-Kai Huang and Chun-Yu Lin

323–336

Flipped Classroom with Problem Based Activities: Exploring Self-regulated Learning in a Programming Language Course
Ünal Çakıroğlu and Mücahit Öztürk

337–349
Designing and Implementing Web-based Scaffolding Tools for Technology-Enhanced Socioscientific Inquiry

Suhkyung Shin*, Thomas A. Brush and Krista D. Glazewski

Department of Instructional Systems Technology, Indiana University Bloomington, Indiana, USA // suhkshin@indiana.edu // tbrush@indiana.edu // glaze@indiana.edu

*Corresponding author

(Submitted September 2, 2015; Revised February 12, 2016; Accepted February 28, 2016)

ABSTRACT

This study explores how web-based scaffolding tools provide instructional support while implementing a socio-scientific inquiry (SSI) unit in a science classroom. This case study focused on how students used web-based scaffolding tools during SSI activities, and how students perceived the SSI unit and the scaffolding tools embedded in the SSI activities. A web-based SSI unit was developed and utilized in a technology-enhanced science classroom, and included three types of embedded annotations (definition, background information, thinking questions). The findings of this study suggest that students benefitted from web-based annotated scaffolding in their individual reading and group discussions. The effective use of hard scaffolding may have contributed positively to students’ engagement in the SSI activities and allowed students to easily explore a variety of context-related hyperlinked resources. In addition, the hard scaffolding tools allowed the teacher greater flexibility to effectively monitor students’ learning progress, evaluate students’ understanding of the topic, and provide guidance as needed through questioning. A holistic and dynamic approach is recommended when teachers and instructional designers consider how web-based hard scaffolding tools might function to assist students’ learning in SSI.

Keywords

Inquiry-based learning, Socioscientific inquiry, Scaffolding, Annotations, Technology-enhanced environments

Introduction

Inquiry-based learning (IBL) has long been considered important for promoting the understanding and retention of concepts, skills and attitudes that are required for solving ill-structured problems. Recently, socioscientific inquiry (SSI) has been emphasized as a curricular model that engages students in scientific topics while considering associated ethical or social issues. In SSI, students develop their understanding of fundamental aspects of science through a question-driven and open-ended process, which includes inquiry activities, planning and managing investigations, and analyzing results (Edelson, Gordin, & Pea, 1999). Specifically, SSI processes include authentic activities designed to motivate learners to acquire and apply new knowledge. Despite the learning benefits of SSI, students often feel frustrated due to a lack of certain assets, such as (1) domain specific knowledge (Bell, Blair, Crawford, & Lederman, 2003), (2) analysis and argumentation skills (Krajcik et al., 1998), (3) the ability to manage information and determine relevance (Hogan, 2002), and (4) the ability to monitor and reflect on their own learning processes (Quintana et al., 2004).

Many researchers have argued that scaffolding provides the framework for assisting students with these challenges. Scaffolds are tools, strategies, and guides that help individual learners to accomplish tasks that are beyond their ability to complete alone (Vygotsky, 1980). Scaffolding can appear in multiple forms depending on the various types of support provided to engage students in an inquiry-based learning activity. Saye and Brush (2002) conceptualized two forms of support: hard and soft. Hard scaffolds are static supports that can be planned in advance in anticipation of potential difficulties with a task. These support structures can be embedded within learning environments to provide students with support while they are actively engaged with a problem (Krajcik et al., 1998; Simons & Klein, 2007). For instance, prompts designed to give definitions or background information for concepts can help students better understand a specific issue during the problem-solving process (Simons & Klein, 2007). In contrast, soft scaffolds are dynamic, situation-specific supports provided by a teacher to help with the learning process. This includes teachers’ clarification of tasks or monitoring of students’ progress (Kim & Hannafin, 2011a), which requires teachers to continuously diagnose learners and provide timely support based on student responses. This type of assistance is generally provided “on-the-fly,” where the teacher monitors students’ progress while engaged in a learning activity and intervenes when support is needed (Saye & Brush, 2002).

Given that hard and soft scaffolds interact in dynamic ways in a classroom context (Kim & Hannafin, 2011a), it is essential to investigate what supports can be provided by scaffolding tools and what supports can be offered by the teacher to optimally facilitate problem-solving among students. For example, as students generate problem
solutions, opportunities to assist students with integrating discrete fragments of evidence into a broader problem context may be incorporated into an inquiry-based unit via additional small-group discussion sessions with the teacher (Saye & Brush, 2004), which would be difficult to provide as a “hard” scaffold.

Scaffolding research typically has focused on certain features and affordances of technology in various settings rather than on the holistic use of scaffolds to support the overall learning experience (Kim & Hannafin, 2011a). Little is known regarding how students experience different types of scaffolding in SSI classrooms. Furthermore, as a relatively new curricular model, few studies have documented the roles of soft and hard scaffolding during SSI activities in classroom practice. Therefore, investigating different types of scaffolding and how they function in different contexts may expand our understanding of how to support and facilitate learning during SSI instruction.

**Literature review**

**Web-based scaffolding tools for inquiry-based learning**

Web-based resources have been widely developed and utilized to support students during IBL activities (Lee & Calandra, 2004; Oliver & Hannafin, 2000). For instance, web-based multimedia resources in multiple formats (i.e., audio, text, visual) can represent various perspectives and enable the presentation of authentic examples, thus promoting learners’ cognitive flexibility (Jacobsen & Spiro, 1995). Despite these potential benefits, it has been reported that learners sometimes feel confused and overwhelmed when utilizing web-based multimedia (Jonassen, 1989; Romiszowski, 1990). In order to relieve learners’ cognitive burden without excluding the potential advantages of web-based learning environments, researchers have highlighted the importance of scaffolding and the benefits of web-based scaffolding tools incorporated into IBL activities (Lee & Calandra, 2004).

Some researchers have reported that web-based scaffolding tools are effective in promoting students’ scientific reasoning skills (Lee & Calandra, 2004; Walker & Zeidler, 2007). For example, Lee and Calandra (2004) reported that annotations embedded in web-based resources encouraged students to access prior knowledge which is essential in understanding contextual information and generating their own explanations during problem-solving. Walker and Zeidler (2007) studied SSI instruction which utilized the Web-based Science Environment (WISE) and found that the environment facilitated students’ exploration of multiple perspectives with various resources. Findings also suggested that engaging in SSI instruction using WISE was not sufficient to promote their understanding of topics as well as acquire scientific skills. Without any guidance in SSI, students produced hasty conclusions or generalizations, or did not make explicit references to a conceptual understanding of the nature of science during classroom debate. However, researchers did find that guiding questions embedded within the WISE environment may have assisted students in recognizing potential bias in information presented to them on-line. This suggests that hard scaffolding tools embedded in web-based instruction may facilitate learner’s scientific inquiry skills in SSI.

As a result of the increased use of IBL, researchers have expanded the classifications of hard scaffolding and guidance about how to integrate hard scaffolds into IBL activities to support students’ learning (Linn, Clark, & Slotta, 2003; Raes, Schellens, Wever, & Vanderhoven, 2012; Williams & Linn, 2002). Hard scaffolding tools include conceptual scaffolds that provide definitions of new terms or web-based resources (Hannafin, Land, & Oliver, 1999), strategic scaffolds that embed expert advice as text-based responses (Simons & Klein, 2007) or video clips to assist students in evaluating alternative approaches to address problems (Pedersen & Liu, 2002), and metacognitive scaffolds that provide evaluation criteria or thinking questions to help students in monitoring and evaluating their progress in completing specific learning activities (Davis & Linn, 2000; Shin & Song, 2015; Wesiak et al., 2014).

Researchers have investigated how learners use web-based scaffolding tools and resources through IBL activities (Belland, 2010; Kim & Hannafin, 2011b). For example, Kim and Hannafin (2011b) explored how 6th graders use peer-, teacher-, and technology-enhanced scaffolds in their classroom during their scientific inquiry activity on WISE, which was used to promote students’ knowledge integration of science topics. Embedded scaffolding, including inquiry maps, hints, and prompts helped learners monitor and reflect on their progress while engaged in inquiry activities. The researchers found that students perceived the embedded scaffolds as useful in helping them focus on important resources to organize evidence needed to support their argumentation. In Belland’s (2010) study, the Connection Log was utilized in a web-based environment, allowing students to respond to prompts and collaborate with peers. Students used the Connection Log to organize information, share their work,
and manage group work throughout the problem-solving process. Results found that scaffolds may assist students in articulating their thoughts and facilitate their thinking processes during problem-solving activities.

**Purpose of the current study**

Although some studies have investigated how scaffolding tools are utilized through IBL, most research has emphasized the effectiveness of hard scaffolding tools in increasing student performance (Kim & Hannafin, 2011a). While measuring students’ achievement has value, there has been relatively little research on how those scaffolds are utilized to support inquiry-based learning strategies such as SSI in a typical classroom. Thus, the purpose of this study was to explore how scaffolding tools can be implemented to support students’ SSI activities in a science classroom. Specifically, this study addressed the following questions:

- How do students use hard scaffolding tools during their SSI activities?
- How do students perceive hard scaffoldings embedded in the SSI unit?
- How do hard scaffolding tools support teacher’s soft scaffolding during SSI activities?

**Method**

**Research design**

This study examined how students utilized and experienced scaffolds integrated into instructional activities, rather than how they evaluated the technology-enhanced classroom itself, by way of an instrumental case study with multiple forms of data (Yin, 2003). Such case studies are used to discover new understandings toward an event, possibly leading toward re-thinking a trend, design, or approach (Marshall & Rossman, 1995).

**Participants**

A 9th grade biology teacher and his 71 students were involved in the present study. Sixty-two students, including 32 males (51.6%) and 30 females (48.4%), completed the reflection survey, and 12 students participated in focus group interviews. The teacher had nine years of experience teaching science and math, and was recruited from among the recipients of a nationwide award program that recognizes expert teachers of inquiry-based approaches with technology.

**Research context**

The context for this study was a high school biology course, comprised of four classes taught by the same teacher, offered in a high school in a rural community in the Midwestern United States. The classes met daily for 50 minutes. This course was selected because this case allows for an in-depth examination of teacher-led development of a science inquiry unit that embedded scaffolding in a technology-enhanced classroom. The course aimed to offer students a unique opportunity to advance their inquiry skills through addressing socioscientific issues.

**Scaffolding design: Socio-Scientific Inquiry Network System**

The teacher developed a unit in a web-based learning environment, the Socio-Scientific Inquiry Network (SSINet), in which students explored authentic socioscientific issues in the classroom. SSINet supports science teachers in their design and implementation of SSI curriculum with web-based curriculum design tools that easily allow them to link to and sequence a wide variety of web-based resources for delivery to students. Teachers can use the “Activity Creator” tool to organize resources and hard scaffolds via a web-based “viewer.”

In this activity, the teacher embedded color-coded hard scaffolding to provide students with guidance. Using the SSINet tools, the teacher was able to embed “definitions (green)” and “background information (blue)” for difficult concepts and “thinking questions (red)” to focus attention on important concepts and issues relevant to the unit (see Figure 1). Students were able to access this information via the activity viewer (see Figure 2).
Teacher-developed inquiry activity

The teacher collaborated with the researchers in developing the SSI unit, which was guided by the following driving question: “When should we use personal genetic information to make decisions?” Students completed a sequence of four SSI activities: Entry event, Jigsaw, Whiteboard, and Culminating activity (see Figure 3).
Figure 3. The SSI Unit activity

Data collection

*Classroom observations*

Two researchers observed classes and recorded field notes during each day of the unit to examine how the teacher used scaffolding to support student progress and how students interacted with resources and tools. Video recordings of the entire classroom and student group activities were also collected.

*Screencasts of student screens*

Eight students’ laptop screens were recorded using QuickTime Player to determine which resources were viewed and used during hands-on activities.

*Focus group interviews*

The authors followed a purposive sampling strategy to select 12 students for focus group interviews (Creswell, 2012). Four groups of students participated focus group interviews. Each group had three students selected by the teacher to maximize diversity of gender, ability, and motivation in the subject area. The interviews explored students’ perceptions of resources and tools that were embedded in the activities. The interviews took place in a classroom, and were audiotaped and transcribed.

*Reflection survey*

A student reflection survey was administered at the end of the unit and examined students’ perceptions about the tools, resources, and inquiry activities. The questionnaire included 10 close-ended and seven open-ended questions. Cronbach’s alpha was used to test the attitude items of the survey and the reliability coefficient was 0.75.
Teacher’s post-unit interview and debriefings

During the unit, daily debriefings were conducted with the teacher, as well as a post-unit interview. Interview/debriefing questions were related to impressions of the class, including the use of resources and tools in his teaching, perceptions about the strengths and weaknesses of this unit, students’ issues or problems, and an assessment of his management of the unit activities.

Procedure

The unit was implemented for eight class sessions. At the beginning of each session, the teacher explained the purpose of the SSI unit and how to access it on the laptops and iPads.

The overarching issue of the unit was introduced with an entry event activity through a case about a woman who had cancer genes. The next six class sessions focused on inquiry activities including: (1) learning about the potential uses of genomic information, (2) examining genetic predetermination and its relation to the environmental influence continuum, (3) investigating when it is legal and illegal to use genetic information, (4) building student opinions, and (5) discussing what laws should govern the use of genetic information.

On the last day of instruction, a reflection survey was administered to students and focus group interviews were conducted. While the class was completing the survey and posttest, students participating in focus group interviews were asked to move to another classroom for their interview. Two researchers conducted the focus group interviews based on a semi-structured protocol and asked six to ten questions based on in-class activities. Students were asked to reflect on unit activities (e.g., How do you think that these activities for the unit were different from other activities you have done in this class?), and provide their perceptions of the web-based tools and resources embedded into the SSI unit (e.g., Do you think the tools and resources which are embedded in the activities were useful to you?). Once they completed the interview, students returned their classroom to complete the remaining group activities and reflection survey. Each group interview lasted 10 to 15 minutes, totalling 50 minutes to complete all four group interviews.

Data analysis

Quantitative and qualitative data were collected from multiple data sources to confirm and interpret conclusions (Creswell, 2012). Screencast data were used as a primary data source in order to measure the frequency of usage of color-coded annotations. The total number of visits and amount of time spent using annotations were calculated. Forty-eight annotations were embedded into the SSI unit: 18 as Definitions, 12 as Background Information, and 18 as Thinking Questions. To investigate students’ perception of the SSI unit, the reflection survey was analyzed with descriptive statistics. Data from students’ and teacher’s interviews, screencasts, and observations were coded by researchers and grouped into conceptual categories related to possible factors affecting student inquiry learning. For the third research question, the teacher’s post-unit interview and observations were analysed and triangulated using comparative analysis to identify reliable themes (Creswell, 2012).

Results

Students’ use of the SSI unit and scaffolding tools

Students frequently used embedded color-coded annotations to explore, find and solve problems using resources that were embedded in the SSI unit. Table 1 summarizes the number of visits to annotations and time spent on each while students interactive the resources that embedded into the SSI unit. The analysis of screen-cast data demonstrated that students accessed annotations that provided thinking questions more often and for longer periods of time than definitions or background information.

Screen-casting data showed that students first skimmed an article by clicking most annotations then later focused on specific annotations when revisiting the resources, particularly the “thinking questions” designed to focus students’ attention on key aspects of a resource. In focus group interviews, four students highlighted benefits of using color-coded annotations, particularly as a starting point for understanding the genetic issues. Most groups
mentioned that they skimmed through resources, went back to annotations and tried to focus on the important aspects of the content, as the following comment suggests:

“…that’s where I really started skimming through it and picking out key points which helped because I could figure out where the key points were by what he highlighted…” (Focus group 1)

Overall, the findings suggested that students perceived annotations as important and used thinking questions to direct their focus.

<table>
<thead>
<tr>
<th>Table 1. Usage of color-coded annotations (n = 8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitions</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>Number of visits</td>
</tr>
<tr>
<td>Time spent*</td>
</tr>
<tr>
<td>(Per annotation)</td>
</tr>
</tbody>
</table>

Note. * = Measured in seconds.

The annotations were also useful when facilitating group discussion. The use of embedded thinking questions helped students to deepen their understanding and analysis during the group activity. In student interviews, one student stated, “…sometimes you click on the red highlighter that gives you questions and when you give it to your partners you discuss them. So, it was kind of interesting, the different questions that they had and how we are giving our opinion on them (Focus group 2).” In addition, data from observations revealed that the activity directions embedded within the activity viewer such as what procedure to follow, what topics to deal with, and what aspects to consider, may have provided strategic scaffolding to assist students as they progressed through the SSI activities. During group activities, the students checked the procedure and key points of group activities by following the prompts with limited guidance from the teacher. In sum, the findings suggested that students perceived the technology-enhanced environments, which facilitated the use of hard scaffolding, as beneficial for helping them focus on the SSI task.

**Students’ perceptions of the scaffolding tools and resources**

**Quantitative findings**

Students perceived that scaffolding tools helped engage them with the SSI unit by allowing them to learn more about the presented problem and facilitating group work. Table 2 summarizes the descriptive statistics for the students’ reflection survey. Results showed that students perceived that the activities helped them learn more about genetics (M = 4.03, SD = .97) and that their experience with group work promoted greater learning (M = 4.21, SD = 1.04). Specifically, students indicated that using mobile devices, such as iPads and laptops, was a positive component in their access to the resources for the unit (M = 4.47, SD = .78).

**Qualitative findings**

Two themes emerged from analysis of qualitative data: (1) authentic resources facilitated student engagement with IBL and may have helped increase their understanding of content, and (2) students struggled with understanding multiple perspectives and providing evidence to support their positions.

The students believed that the resources and materials embedded in the SSI unit were useful and helpful in terms of increasing their understanding and strengthening their grasp of concepts. Specifically, four groups mentioned in the focus group interviews that various materials, such as video clips, annotated articles, and other media, were authentic and context-based resources related to their real life, which helped them connect with and understand the material:

“My favorite thing was reading the articles. I liked learning about other people’s lives, how genome sequencing affected them…” (Focus group 3)

“I just liked like reading the articles, and watching some of the videos. It helped me understand it better.” (Focus group 2)

However, observation and interview data indicated that some students found it difficult to support their position using evidence provided in the culminating activity. Specifically, when asked to identify the most difficult
activity in the SSI unit, most of the students mentioned the challenges involved in developing their group presentations. During the focus group interview, one student discussed the challenges associated with this activity, particularly with respect to supporting her opinion with evidence: “Sometimes just agreeing with the rest of the group because you know somebody would put down something and you would be you didn’t all really decide on that. And sometimes coming with the reasons, it’s like you know what you want to say but you don’t know how to back it up.” (Focus group 4)

In addition, observation data revealed that some of the groups failed to provide evidence to support their position on the use of personal genetic information. The groups were able to discuss their chosen position on the issue, but the evidence they used to support their position was weak or missing altogether.

Table 2. Students’ reflection survey (n = 62)

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I enjoy science class.</td>
<td>3.90</td>
<td>1.04</td>
</tr>
<tr>
<td>I do well in science class.</td>
<td>4.19</td>
<td>.90</td>
</tr>
<tr>
<td>I would like to learn more about how genetic information is used to make decisions.</td>
<td>3.23</td>
<td>1.16</td>
</tr>
<tr>
<td>I enjoyed using the iPad/laptop to access the resources for this unit.</td>
<td>4.47</td>
<td>.78</td>
</tr>
<tr>
<td>The activities I completed in this unit helped me learn more about genetics.</td>
<td>4.03</td>
<td>.87</td>
</tr>
<tr>
<td>I enjoyed the group work I completed in this unit.</td>
<td>4.30</td>
<td>.98</td>
</tr>
<tr>
<td>I think working in a group helped me do better on the activities for this unit.</td>
<td>4.21</td>
<td>1.04</td>
</tr>
<tr>
<td>I enjoyed completing the final project for the unit.</td>
<td>3.82</td>
<td>.97</td>
</tr>
<tr>
<td>I wish the teacher had provided more guidance to my group and myself during the unit.</td>
<td>2.22</td>
<td>1.00</td>
</tr>
<tr>
<td>I would like to study other science topics the same way we studied genetics in this unit.</td>
<td>4.14</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Note. 1 = strongly disagree, 5 = strongly agree.

The Role of hard scaffolding tools and soft scaffolding

Supporting self-directed learning

Analysis of student data obtained from focus group interviews and open-ended survey questions revealed that scaffolding tools and resources may have enabled students to conduct their own research more independently. For example, comments made by students during focus group interviews suggested that by clicking the hyperlinks, the students easily accessed necessary information and resources to successfully complete specific activities. One student stated, “it was really organized, and you could easily find things with simple instructions, you can find stuff, it was all set up for you, so you didn’t have to spend half the class finding it (Focus group 1).”

In addition, interview data suggested that hard scaffolds embedded in the activities may have provided more opportunities for independent individual learning. Six students (50%) mentioned in the interviews that utilizing the hard scaffolding allowed them to better manage their learning. For instance, one student explained: “Something that I like personally was you have a lot more freedom … and you can work on your own pace as long as you got it done. You didn’t have to try rushing through it or miss things but you also didn’t have to get through, stop, and wait for half an hour until the rest of the class catches up (Focus group 4).”

The teacher also mentioned in the post-unit interview that hard scaffolds (such as embedded annotations) were useful for his students in terms of exploring the learning content without his direct support: “… They (students) were able to click on certain words, find out the definitions of the words without me having to explain them, or using a cumbersome dictionary tool. I thought it was useful for that.”

Although hard scaffolding was provided while students were in the process of discussing and developing their presentations, observation of the class suggested that some students still struggled with solving problems. In such cases, soft scaffolding by the teacher may have supported the students by providing modeling and questioning. While conducting the culminating activity, the students experienced challenges to developing their presentation, and the teacher provided specific examples and asked questions to shape their thinking and guide them through the activities. Specifically, survey data showed that the teacher provided sufficient soft scaffolding, which might influence students’ satisfaction with their learning experience (see Table 1). In response to the question “I wish the teacher had provided more guidance to myself and my group during the unit,” only 10% of students indicated that they felt they needed more guidance. In response to the question, “I would like to study other science topics the same way we studied genetics in this unit,” over 75% of students indicated that they enjoyed the experience.
Enabling the provision of timely soft scaffolding

Providing hard scaffolding may have allowed the teacher to assist other students who needed help. In a focus group interview, one student noted: “If he thinks you understand it, then he’ll just let you do it by yourself, he spends more time with kids that need help with…things (Focus group 3).” The data from observations, and the teacher’s debriefings and interview support this student’s assertion. While observing the class, researchers found that the teacher was able to monitor students’ progress through the activities, and assess their level of understanding of the content. The teacher moved through the classroom and periodically asked questions of students to both facilitate their understanding and to determine which students may be having difficulty with the content.

The teacher’s comments in debriefings and interviews also suggested that hard scaffolding enabled him to secure time for providing soft scaffolding. As he stated, “… [the annotations] allowed me to go around and ask questions as they are reading, because I could kind of see where a student was at, and ask a pertinent question based on that.” The teacher also noted that he was able to have more interaction with his students and assess their progress, which in turn allowed him to both adjust the support he provided to some students while in turn giving other students greater independence.

Discussion

The purpose of this study was to examine how learners used the resources and scaffolding tools embedded in an SSI unit to support their learning, and how hard scaffolding tools could support teacher soft scaffolding. The first research question examined students’ use of resources and hard scaffolding during their SSI activities. In this study, students benefitted from color-coded annotated scaffoldings in their individual reading and group discussion, specifically as they were identifying SSI issues, exploring background information, and researching evidence. We found that students considered the annotations as supports that the teacher intentionally provided for additional information on reading materials. Analysis of data suggested that students used the annotations as a starting point for helping them better understanding the content and assisting them with their reading comprehension. In addition, based on students’ interviews and observations, we identified three benefits of web-based color-coded annotations that may play an important role in promoting greater student engagement in inquiry activities: (1) directing attention to important information, (2) providing structured guidance that enhances learners’ interpretation and analysis of problems, and (3) facilitating questions that highlights critical aspects of the problem. As previous studies have suggested, different hard scaffoldings, which were provided in this study as conceptual, strategic and metacognitive scaffoldings, can be utilized to alleviate learners’ difficulties and facilitate the problem solving process (Shin & Song, 2015; Simons & Klein, 2007). However, the present findings attempt to address the limitations of past studies that investigated the effects of scaffolding on student achievement by identifying how students recognize and use scaffolds to acquire content knowledge or background information during IBL activities.

Another important finding was that students preferred to use strategic scaffolds (e.g., thinking questions, activity guides) while investigating content and participating in activities and group discussion. As a strategic scaffold, thinking questions may help students develop and refine their reflective thinking because question prompts may serve as a cue or guide to focus their attention and continually monitor their learning as they explore the content (Davis & Linn, 2000; Ge & Land, 2003). In SSI, learners deal with problems that accommodate multiple perspectives and incorporate theory or principles in evaluating and supporting their arguments. Research suggests that learners struggle to conduct disciplined SSI with substantial learning loads when simultaneously managing the necessary reasoning skills and new content, and strategic scaffolding may relieve their cognitive burden and promote their problem-solving processes (Walker & Zeidler, 2007).

The second research question investigated learners’ perceptions of the SSI unit and scaffolding tools. The findings suggest that the effective use of hard scaffolding may have contributed positively to students’ engagement in the SSI activities, especially through the activity viewer which allowed students to easily explore a variety of context-related hyperlinked resources. In SSINet, the activity viewer presented resources with embedded hyperlinks on the left side, which enabled students to explore the resources on the right side of panel by clicking the hyperlinks. This structure also facilitates the navigation of other resources (i.e., review the guidelines of activities, revisit resources from the left panel). This SSINet tool feature may promote learners’ cognitive abilities to investigate essential information without eliminating the benefits of non-liner features found in hypermedia learning environments. In addition, in this study, the students indicated that the resources related to their own life experience were helpful for understanding multiple perspectives and increasing their
understanding of actual social problems. These different interactive scaffolds may positively impact students’ satisfaction with the SSI unit and acquisition of content knowledge. While the rich and numerous resources seemed to play a positive role in the students’ initial engagement, these resources and scaffolds may also have sustained effects as students build and contextualize new knowledge in meaningful ways (Schank, 1999).

However, students encountered challenges while conducting the culminating group activities, suggesting that additional hard scaffolds may be needed to support their learning throughout the entire process. Specifically, student feedback indicated that learners sought additional annotations, including examples or expert explanations, for help with representing the central issue from multiple perspectives and modelling problem-solving processes such as developing claims and evidence. Students’ difficulties with linking evidence to specific problem contexts might be due to their lack of experience in making arguments and limited opportunities to develop and support knowledge claims with evidence (Kim & Hannafin, 2011b; Krajcik et al., 1998; Zeidler, Walker, Ackett, & Simmons, 2002). For example, Kim and Hannafin (2011b) found that 6th grade students failed to examine evidence during a science inquiry activity. They reported that learners’ difficulty in generating their arguments with supporting evidence was influenced by their limited inquiry strategies and understanding. The results of this study suggest that additional hard scaffolding, such as strategic scaffolds that guide students while building arguments, might be needed to facilitate more effective decision-making processes among student groups while developing their group presentations.

The third research question focused on the role of soft and hard scaffolding. The findings suggest that hard scaffolding may play a significant role in increasing teachers’ ability to provide timely soft scaffolding to students. While interacting with the resources and scaffolding tools embedded in the SSI unit, some students were able to independently identify problems, investigate resources and content, and construct their own arguments. This in turn allowed the teacher greater flexibility to effectively monitor students’ learning progress, evaluate their level of the understanding, and provide guidance as needed through questioning.

Since individual learners frame their arguments around their own experiences and prior knowledge (Kim & Hannafin, 2011b; von Aufschnaiter, Erduran, Osborne & Simons, 2008), they may need specific and differential support to complete various tasks required of them during an SSI unit. However, a teacher may not be able to provide appropriate guidance to facilitate problem-solving processes to individual students in K-12 classroom environments because of limited time and resources (Hmelo-Silver, 2004). Similar to previous studies on hard scaffolding in classroom environments (Saye & Brush, 2001; Simons & Klein, 2007), the results of this study suggest that providing differential hard scaffolds can create more time for teachers to provide soft scaffolds when necessary. In this study, hard scaffolds helped guide some students through the process of solving socioscientific problems by themselves, which allowed the teacher to prioritize other students’ needs for additional support. The teacher was able to provide timely soft scaffolding by questioning and modelling while monitoring and evaluating all of his students’ progress through the SSI activities.

**Implications and suggestions for future research**

Although this study has allowed us to explore the role of hard and soft scaffoldings in a typical science classroom environment, the findings cannot be generalized, and should be articulated and expanded with further research in different settings. Nonetheless, the results of this study do contribute a clearer and more nuanced perspective of teachers’ and students’ experiences with scaffolding in technology-enhanced classroom environments, and help to refine conceptions of hard and soft scaffolding related to inquiry-based learning activities. First, hard scaffolds embedded in the SSI unit might support learners in overcoming barriers to dealing with ill-structured problems, which are central to SSI activities. Given that students focused more on the annotated elements, instructional designers should consider carefully how students perceive scaffolding tools and how and what condition they used it. Practically, utilizing scaffolding tools interacted with complex and dynamic situations in an actual classroom. Considering that students’ prior knowledge and experiences may affect their use of scaffolding tools, further research is needed to investigate learners who experienced challenges during the inquiry processes in order to gain additional insight regarding improved scaffolding design to support those learners.

Secondly, providing authentic and relevant resources that directly link the problem to authentic situations may have been helpful in terms of building students’ background knowledge and provoking their initial motivation. Most students reported that the resources and scaffolding embedded in the SSI unit were interesting and meaningful due to their realistic relationship to real-world problems. In order to enhance the impact that
scaffolding has on learners’ engagement, it is crucial to provide authentic resources and tasks that students can meaningfully relate to their own experiences.

Finally, it is important to understand and clarify the teacher’s role during disciplined inquiry in technology-enhanced classroom environments. Although the hard scaffolds embedded in the unit activities did assist learners’ inquiry activities, the teachers’ soft scaffolding, in the form of questioning or modelling, remains essential for monitoring student progress and promoting greater understanding. However, results of this study did suggest that the hard scaffolds (particularly the annotations and the linked resources) assisted the teacher with targeting additional soft scaffolding to students who were struggling with specific SSI activities. To better support IBL activities in technology-enhanced environments, researchers should further explore the relationship between hard and soft scaffolding, and consider how to best align scaffolding supports with different elements of the learning environment. To accomplish this, further research is needed to investigate in more depth the relationships and outcomes of employing different types of soft scaffolding in tandem with hard scaffolding in classroom environments.

References


Are They Thinking Differently: A Cross-Cultural Study on the Relationship of Thinking Styles and Emerging Roles in Computer-Supported Collaborative Learning

Xiaoqing Gu¹, Huawen Wang¹, and Jon Mason²

¹East China Normal University, China // ²Charles Darwin University, Australia // xgu@ses.ecnu.edu.cn // bigwanghuawen@163.com // jon.mason@cdu.edu.au

*Corresponding author

(Submitted September 6, 2015; Revised December 19, 2015; Accepted January 21, 2016)

ABSTRACT
Numerous studies have recognized collaboration as an effective way of learning. When collaboration involves students from different cultural backgrounds, a question arises: Will cultural differences influence the manner in which roles are adopted within collaborative learning? In this study, a correlation analysis was used to explore the relationship between cultural factors and emerging roles among collaborating students from two universities in different countries (China and USA). The cultural factors that might hypothetically affect their collaboration were approximated to thinking styles by using Sternberg’s thinking styles inventory. The roles that students adopted according to preferences were coded with an adapted coding scheme. The results indicate a significant relationship between student thinking styles and the adopted roles of students. This finding implies that cultural factors, exhibited as thinking styles, could explain the emerging roles that are adopted in Computer-Supported Collaborative Learning (CSCL). The results could guide teachers in assigning appropriate roles to students with different backgrounds to improve the efficiency of collaboration during cross-cultural CSCL.

Keywords
Cross-culture, Computer-Supported Collaborative Learning, Thinking styles, Emerging roles, Correlation analysis

Introduction

Discussions and debates have focused on the effects of increasing globalization on education (e.g., Zajda, 2015; Marginson, 2011). The statement, “[t]he global dimension is fast-moving and fluid” (Marginson, 2011, p. 11) indicates a work-in-progress that deals with emergent issues constantly playing out. Worldwide participation rates in higher education have clearly risen in the last decade (OECD, 2012), and this increase is widely interpreted as an indicator of socio-economic advance. Moreover, in several cases, many classrooms and courses are experiencing an increasingly internationalized student population. This scenario has raised a number of issues and opportunities, such as political consequences in certain contexts as well as an increasing need for educators to deal with cross-cultural issues in teaching and learning (Nguyen, Elliott, Terlouw, & Pilot, 2009; Vatrapu, 2008) and specifically with learning online (Nafukho, 2014).

Evidence from psychological research has identified individual learning differences, which emerged from cultural factors. Specifically, these differences have been described at the level of the individual in terms of motivation and cognitive process (Millar, Serbun, Vadalia, & Gutchess, 2013; Han, 2010), thinking styles (Lun, Fischer, & Ward, 2010), and learning styles (Joy & Kolb, 2009). These differences manifest as individual behavior, and much of the literature discusses these factors in terms of cultural differences. Likewise, since the emergence of Sternberg’s work on thinking styles over two decades ago, much research has been undertaken concerning the relationship between thinking styles and other variables, such as “cultural differences with thinking style” (Han, 2010; Lun, Fischer & Ward, 2010; Millar et al., 2013). In engaging in international collaboration, these issues come to the foreground, as this study reveals.

In reporting a study focused upon cross-cultural factors that might affect role adoption within Computer-Supported Collaborative Learning (CSCL), this study considers a hypothesis that people from different cultures think differently and will therefore behave in different ways by adopting different roles in a cross-cultural CSCL context. We first review related literature on cross-cultural studies on learning, thinking styles, and role-playing in CSCL contexts. We then report on a study designed specifically to investigate whether a correlation could be determined between student cultural contexts and the roles they adopt in CSCL activities. We assume a broad interpretation of CSCL throughout the study, although a number of the activities described could arguably be distinguished as distributed problem-based learning (DPBL). Our rationale for choosing CSCL as the primary...
context for this study is that collaboration indicates various roles that need to be performed in addressing prescribed tasks.

**Literature review**

**Cross-cultural studies in learning and CSCL**

Culture is a construct commonly used across a range of discourses that include anthropology, sociology, cross-cultural studies, and organizational development. At an abstract level, the components of culture comprise shared norms, values, symbols, and beliefs (Hofstede, 1980; 2011; Triandis, 1995; Zhong, 2010). Culture becomes an important construct because human societies and groups conduct their affairs with different conventions and protocols. Hence, this situation provides both opportunities and constraints for mediated interaction.

Learning and teaching are culturally dependent phenomena, and much research on learning consider cultural factors, with extensive evidence showing the differences produced by cultural factors (Millar et al., 2013; Han, 2010; Lun, Fischer, & Ward, 2010). While several studies demonstrate the efficacy of generalization (Van de Vijver & Poortinga, 1982), further analysis that considers a range of contexts reveals significant complexity. For example, while cultural dispositions, such as respect for hierarchy, can be viewed as predicting aspects of organizational harmony over individual needs, other factors, such as overall cultural composition of workplace teams and whether frequent communication promoted early in a team’s lifecycle, are as significant in determining performance (Swigger, Alpasan, Brazile, & Monticino, 2004). However, in other studies, Popov, Biemans, Brinkman, Kuznetsov, and Mulder (2013) found that “collaboration in culturally mixed groups is less than optimal and may require extra facilitation” (p. 36).

Zhong (2010) shows that the “cultural influence in CSCL is multi-faceted” (p. 168) and that “cultural [dimension] has been found as an important factor in affecting the collaborative process, directly or indirectly, and the learning outcomes in CSCL” (p. 178). Her thesis explicitly investigates cultural influence in CSCL and explores the expression of cultural values as individual characteristics and the ways that these values could determine technology acceptance and use.

In the case of education, the dimension of power distance manifests as a large distance in teacher-centered education and as a smaller distance in more student-centered education (Hofstede, 2011, p. 9). However, this study considers the dimension of individualism versus collectivism to be more significant. Culture is an important factor in shaping and predicting the action and response of individuals. Because many Asian cultures are often characterized in terms of supporting collectivist goals (Miyahara, Kim, Shin & Yoon, 1998), these cultural dispositions can affect the perception of CSCL tasks (Zhong, 2010).

**Thinking styles in cross-cultural studies**

Cross-cultural studies reveal that cultural factors affect different behaviors and produce culturally unique thinking styles (Han, 2010; Millar et al., 2013; Varnum, Grossmann, Kitayama, & Nisbett, 2010). These differences in thinking styles are apparent in cognitive process, such as perceptual and attentional processing by a social psychologist (Han, 2010). However, these differences are apparent in terms of cognitive habit according to cultural psychologists, and reveal a more analytic and holistic pattern of thinking and perception (Varnum et al., 2010).

In the seminal work on thinking styles of Sternberg (1990), cultural construct is introduced as “mental self-government” to highlight the relationship between individual abilities and preferences and its effect on teaching and learning (p. 369).

Thinking styles are as important as, and arguably more important than abilities, no matter how broadly abilities are defined. Thus, constructs of social, practical, and emotional intelligence, or of multiple intelligences, expand our notions of what people can do, but the construct of style expands our notion of what people prefer to do or how they capitalize on the abilities they have. When your profile of thinking styles is a good match to an environment, you thrive. When the match is bad match, one suffers. Different levels of schooling and subject areas reward different styles as well (Sternberg, 1999, p. x).
In using the notion of “government,” Sternberg describes its various forms (monarchic, hierarchical, oligarchic, and anarchic) as well as its key functions (executive, legislative, and judicial). Distinguishing preference from ability has provided an important theoretical lens through which subsequent research has focused concerning the biases or preferences in applying thinking styles (Zhang & Sternberg, 2000; Varum et al., 2010).

Within sociocultural context, thinking styles are subject to cultural differences using broad descriptors, such as analytic versus holistic (Zhu & Han, 2008). Therefore, in this study, approximating cultural factors to thinking styles, while using Sternberg’s thinking styles inventory made sense. Applying Sternberg’s distinctions of preference versus ability within the context of CSCL requires close examination of roles within CSCL.

Roles in CSCL

The concept of role has been widely used in CSCL and is defined as the stated functions or responsibilities that guide individual behavior and regulate group interaction (Hare, 1994). Generally, two perspectives of roles are found in the practice of CSCL (Strijbos, & Weinberger, 2010). First, scripted roles facilitate the collaborative learning process by prescribing and structuring roles and activities to learners. Second, emerging roles are developed spontaneously by participants during their collaborative learning process.

Structured role design has been shown to be an effective way to engage students in collaborative learning activities in both face-to-face and online contexts (Gu, Shao, Guo, & Lim, 2015; Hoadley, 2010; Hou, 2012; Yeh, 2010; Wishart, Oades, & Morris, 2007). By assigning a student with a particular role in the CSCL process, issues of interdependence, individual accountability, and cognitive engagement in collaborative learning are addressed (De Wever, Van Keer, Schellens, & Valcke, 2010; Strijbos & Weinberger, 2010). However, studies on role structuring rarely consider characteristics of individual student preferences as well as cultural differences between students when assigning them to different roles.

Research on emerging roles has focused on the analysis of de facto roles that students produced, self-regulated, and developed (Strijbos & Weinberger, 2010). The analysis of emerging roles is valuable for understanding the CSCL process, including the dynamics of CSCL, individual contribution, and patterns of interactions with each other (Jahnke, 2010; Sarmiento & Shumar, 2010), which are important knowledge for effectively facilitating CSCL.

As mentioned earlier, cross-cultural factors have significantly influenced social engagement in collaboration (Van de Vijver & Poortinga, 1982; Zhong, 2010). To document the cultural dimensions of CSCL by structuring roles appropriately, studies on the possible relationship between cultural differences and their roles in CSCL are needed.

This study explores the relationship between cultural factors and emerging roles among collaborating students from two universities in different countries (China and USA). The cultural factors that might hypothetically affect collaboration were approximated to thinking styles by using Sternberg’s (1990; 1999) thinking styles inventory, while the roles that the students adopted were coded with an adapted coding scheme. The research questions adopted as our foci are as follows:
- Are there significant differences among the roles they played in collaboration?
- Are the differences of their roles played in collaboration significantly related to their thinking styles?

The following sections describe the methodology of this study, including the participants, research design, coding scheme adopted, and approach to data collection and analysis.

Methodology

Participants

The participants in this study were 27 Chinese graduate students (7 males and 20 females, aged 23 to 25 years old) majoring in Educational Technology, and 32 American graduate students (12 males and 20 females, mostly in their 30s) majoring in Curriculum during the 2013–2014 Northern hemisphere fall semester. The Chinese were all full-time graduate students with no work experience, while the Americans were pre-service science teachers and all part-time graduate students about to become PhD candidates. Because of their majors, the Chinese students had skills in software development, while American students had skills in instructional design.
This situation provided an optimum combination of different skills that the collaboration expected. All students had some knowledge of and experience with CSCL, which provided good foundation for the success of the project. All Chinese students studied and learned English for more than eight years as a second language and had passed the College English Test 6, which meant that English could be used during the collaboration. Students were divided into 13 groups, with each group comprising two Chinese and two or three Americans.

Research design

The context of the study builds upon a well-established collaborative relationship between two universities, one from US and the other from China. Each academic year, a cohort of graduate students from the universities work together on projects within their respective courses to design and develop educational games for K12 students. The teachers introduced the targeted users, demand and assessment criteria, basic design and technological knowledge, and the timeline of the project. However, the entire project design and development of the games is intrinsic to the project. A key feature is that the course combined elements of collaboration among different roles and instructional design.

At the beginning of the project, students contacted group members through email and added them as contacts on Skype, which was the chosen platform for communications during the following semester. Communication among group members across the countries was conducted in English. However, Chinese was used when the conversation involved only Chinese group members.

The task of each group was to design and develop collaboratively an educational game. Each group was expected to plan their task schedule towards this goal. Students were required to engage in planning by discussing the topic once or twice on each week and the task that must be finished for the following week. Moreover, these meetings assured that future meetings were scheduled. The duration of each meeting was scheduled for one and a half hours, although the actual duration depended on the efficacy of the discussion.

To achieve the learning goal, students collaborated using Skype and provided input into the requirements, design, development, debugging, and evaluation of the game software. Each communication episode was regarded as an instance of CSCL. Each group was required to discuss any issues and come to an agreement in one session, such as the theme of the educational game they would design, what learning strategies they were going to use, or what the game should look like. Each group worked independently on their game design activity, and their output was the game they designed and developed together. During the whole semester, the instructors only provided technical support for this collaborative project. No further intervention was done during the process of the students engaging in CSCL.

Coding emerging roles

In this study, we observed the roles that the students adopted in the collaboration, and coded these observations by using a CSCL role-coding scheme. This coding scheme illustrates six steps that normally involve resolving a conflict, finishing a discussion task, or solving a problem. The role structure we used to approximate the emerging roles of the students is presented in Table 1.

The starters are responsible for setting the timeline of the whole project, kicking off the discussion, putting forward a preliminary analysis of the task, and motivating discussion when necessary. The supporters assist peers by giving positive feedback to their inputs with supportive evidence. The arguers will elaborate upon or add further explanation and analysis to statements made by peers according to logical and critical thinking. The questioners will focus on raising queries to clarify statements or solutions. The challengers provide strong negative feedback and ask critical questions that might reveal assumptions. The timers will coordinate and moderate the speed of discussion, set the tone for discussion, and make a summary at the end of the discussion (Gu et al., 2015).

Students were required to record the Skype video of each CSCL session and save the conversational transcript as text. At the end of the project, all videos and transcripts were sent to the researcher as raw data for coding.

Two research assistants were assigned to handle coding of the emerging roles of the students. Coding was based on Skype videos, interview transcriptions, text messages, or emails as necessary. The code was scaled at 1 to 9.
with 1 indicating that the student had never enacted this role and 9 indicating complete engagement with this role. The higher the digit recorded, the higher the level of the role-play experience. Table 2 presents the coding example.

**Table 1. CSCL roles**

<table>
<thead>
<tr>
<th>Roles</th>
<th>Function description</th>
<th>Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starter</td>
<td>To kick off the discussion, put forward a preliminary analysis of the task, add new points for peers to build upon, and give new impulses when discussion slacks off; To make a summary in a certain time, and promote development of the collaboration</td>
<td>To begin with, I think We can first make sure Let’s come to the point Let me say more about that From the story, it is clear that Let us solve the problems, now</td>
</tr>
<tr>
<td>Supporter</td>
<td>To support ideas of peers by making positive feedback with reasonable evidence</td>
<td>I agree because That’s right I can see what you are saying An example is I have read that</td>
</tr>
<tr>
<td>Arguer</td>
<td>To make further explanation/reasoning to the statement/idea made by peers, with logical and critical thinking</td>
<td>That is valid if I think both are right because To summarize From the discussion, we can see So, what you mean is It sounds great since</td>
</tr>
<tr>
<td>Questioner</td>
<td>To raise questions and not doubt, from statements of peers; To promote discussions as planned</td>
<td>Why is it? What do you mean when you say? Can you say more on that? Is there another way of looking at it? Where did you read/hear that?</td>
</tr>
<tr>
<td>Challenger</td>
<td>To challenge statements/ideas and give negative feedback by asking critical questions and probing into their opinions; To lead teams to consider the problem critically and logically</td>
<td>I disagree/am not so sure because Why do you say that? Please give a reason Is there any evidence? But, can we trust that? An argument against that is/ Another view might be I think something different Is it the case that?</td>
</tr>
<tr>
<td>Timer</td>
<td>To coordinate and moderate the speed of discussion, set the tone for discussion, respond to individual posts, and prompt individuals or the group to pursue ideas</td>
<td>Would you please? Can we? Sorry Ok. Let’s move on Would you please?</td>
</tr>
</tbody>
</table>

**Table 2. Role coding examples**

<table>
<thead>
<tr>
<th></th>
<th>Researcher 1</th>
<th></th>
<th>Researcher 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sta</td>
<td>Sup</td>
<td>Arg</td>
<td>Que</td>
</tr>
<tr>
<td>Student 1</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Student 2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

Prior to coding the data, the two research assistants received approximately two hours of training in understanding the coding scheme and elaborating on the coding process. After the two-hour training, sample data were used for practice. The final coding results were pooled into MS Excel. Inter-rater reliability (IRR) was calculated by using Cohen’s Kappa. Table 3 presents the IRRs.

We can see from Table 3 that six IRRs are all in (0.7, 0.8), which indicates high level of reliability of the two raters. Moreover, the coding results could be used for further correlation analysis.
Table 3. Inter-rater reliability calculation

<table>
<thead>
<tr>
<th>Role</th>
<th>Starter</th>
<th>Supporter</th>
<th>Arguer</th>
<th>Challenger</th>
<th>Questioner</th>
<th>Timer</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRR</td>
<td>0.842</td>
<td>0.805</td>
<td>0.801</td>
<td>0.760</td>
<td>0.737</td>
<td>0.886</td>
</tr>
</tbody>
</table>

Group interviews were carried out to refine the coding decisions. However, because of difference in region and time, only Chinese group members were interviewed to obtain better understanding of the role-play in CSCL. The researchers first contacted one of the Chinese group members, explained the interview purpose, and promised to respect personal privacy. After negotiating with their own group members for a couple of days, all 13 groups took part in the interview. The interviews were done with each group at the time and place convenient to them. With permission, the interviews were recorded for later analysis and transcription. During the interview, each student was asked to describe in detail the team collaboration and contribution in CSCL. At the end of the interview, the researchers described the six roles in Table 1, and asked the students to evaluate the roles that each student adopted, including American members. This task was done blind to what other the group members thought, and the results of the role evaluation were used as the source of triangulating the coding results.

Thinking skills data

Many scales can be used to measure thinking styles, which could explain the difference in individual academic performance (Grigorenko & Sternberg, 1995). The thinking styles inventory of Sternberg and Wagner (1991) can measure student thinking styles from 13 dimensions, including 104 items, with a seven-point Likert scale (1-Not at all and 7-Extremely well). This thinking styles inventory has been used in different grades by many researchers and the results showed high reliability and validity. However, the question list is too long, which could agitate participants answering the questionnaire. Such agitation have been considered to affect the reality of the measured results (Lin, Qin, & Chen, 2008). To acquire answers with better accuracy from the participants, Chinese researcher Lin and colleagues (2008) adapted and validated the original version to a shorter one, which has only 64 questions. In this study, we used this short version as the instrument. The measurement was carried out at the beginning of the semester and the results were exported as a spreadsheet. Because the short version has a different number of questions in each dimension, the results were transferred to a hundred-mark system for comparison and calculation. A higher score indicates stronger thinking style on a certain dimension.

Findings and analysis

Results and analysis of thinking styles

After encoding the questionnaire data into a spreadsheet, the average scores of the thinking styles of students from the different countries were calculated from the 13 categories. ANOVA was employed to analyze the differences in thinking style dimensions. The results are shown in Figure 1 and the significant differences are listed in Table 4.

Table 4. ANOVA test of thinking styles

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judicial</td>
<td>5.296</td>
<td>1</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Liberal</td>
<td>17.424</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Hierarchical</td>
<td>18.208</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Oligarchic</td>
<td>7.258</td>
<td>1</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Anarchic</td>
<td>14.998</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>52</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1 shows that visually, Chinese students achieved higher scores in oligarchic, global, executive, conservative, and local styles. By contrast, they had lower scores in liberal, hierarchic, anarchic, judicial, and legislative in comparison with American students. Combined with the ANOVA test results in Table 4, among the 13 thinking styles, significant differences are observed in liberal (\( p = .000 \)), hierarchical (\( p = .000 \)), anarchic (\( p = .000 \)), oligarchic (\( p = .010 \)), and judicial (\( p = .025 \)) thinking styles between Chinese and American students. Executive, global, conservative, and oligarchic are not significant, and local and monarchical almost have no disparity. According to Sternberg’s theory of mental self-government:

“Individuals with liberal style enjoy engaging in tasks that involve novelty and ambiguity” (Zhang & Sternberg, 2000, p. 475);

“The hierarchic form allows for multiple goals, each of which may have a different priority” (Sternberg, 1990, p. 369);

“For individuals who have anarchic thinking styles, rules, procedures, and guidelines are anathema” (Sternberg, 1990, p. 369);

“The oligarchic form allows for multiple goals, all of which are equally important” (Sternberg, 1990, p. 369); and,

“The judicial mind is concerned with judging, evaluating, and comparing” (Sternberg, 1990, p. 367).

Therefore, we make the following deductions:

- American students prefer new ideas and new methods or solutions to solve problems that involve novelty and ambiguity. A significant difference (\( p = .000 \)) can be observed in the liberal thinking style with Americans 82.20: Chinese 68.47.
- American students prefer to distribute attention to several tasks prioritized according to their value to the individual in achieving his or her goals. A significant difference (\( p = .000 \)) can be observed in the hierarchical thinking style with Americans 81.65: Chinese 68.78.
- American students enjoy working on tasks that allow flexibility as to what, where, when, and how one works, while tending to eschew systems of almost any kind. A significant difference (\( p = .000 \)) is shown by anarchic thinking style with Americans 78.57: Chinese 68.47.
• Chinese students are more likely to work on multiple activities in the service of multiple objectives, but do not enjoy setting priorities. A significant difference ($p = .010$) is indicated in oligarchic thinking style with Chinese 56.70: Americans 67.72.
• American students are more willing to focus on evaluating the products of other groups. A significant difference ($p = .025$) is shown in judicial thinking style with Americans 75.60: Chinese 67.72.

Results and analysis of roles

The average scores of the roles played were calculated according to the coding results. The results are shown in Figure 2, and ANOVA test results of the significant differences are listed in Table 5.

We can see from Figure 2 and Table 5 that Chinese and American students were playing their roles very differently in questioner ($p = .000$), challenger ($p = .002$), arguer ($p = .016$), and starter ($p = .041$). ANOVA test results suggested that Chinese students were more likely to play roles of arguer, questioner, and challenger, while most American students preferred roles as starter.

<table>
<thead>
<tr>
<th>Table 5. ANOVA test results of role-play</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Starter</td>
</tr>
<tr>
<td>Arguer</td>
</tr>
<tr>
<td>Questioner</td>
</tr>
<tr>
<td>Challenger</td>
</tr>
</tbody>
</table>

During the interview, the researcher asked the students to describe the details of collaborative processes to find out how the members played the roles. The following quotes that echo the ANOVA test results are representative (names mentioned are aliases):

“At the beginning of the project, it’s the American friends, mainly York, who calls us on the Internet. He sets the online meeting time and sends us emails or messages to know the task process.” (American students incline to be starters)

“They are so fluent in English that I can’t follow up. I usually tell my Chinese partner my opinions and he would translate for me because his English was better. I mainly focus on the task assigned to me, such as add a scenario in the game, and ask necessary questions.” (Chinese students like to be questioners)

“I don’t think what they design is an education game, but a courseware. However, our goal is to design and develop an education game. Therefore, we negotiated and I proposed that the content was still there but wrapped in game clothes.” (Chinese students were inclined to be challengers and arguers)

“They may not know us well. We are not professional programmers to develop the entire design of the game. Part of the designed game was too hard to develop. I advised to change the design to an easy one and they agreed. Another example was that the background music they use was noisy. I recommended another one and they think it’s a great idea.” (Chinese students were inclined to be challengers and arguers)

“They don’t have much reject voice. They usually say ‘It looks great!, ‘You have done great job’ and so on.” (Americans do not like to be challengers and questioners)
Figure 2. Average scores of role-play

From Figure 2 and the interview, American students are outgoing, enthusiastic, nice to group members, and passive on organizing the collaborative learning. Chinese students, however, tended towards being a little shy. Although communication in English is not a problem in terms of understanding, the Chinese lacked confidence and do not speak much. They would rather focus on their own assignment and express ideas when necessary. Therefore, Chinese students naturally tended toward the roles of arguer, questioner, and challenger.

Correlation analysis of thinking styles and roles

After first identifying the thinking styles and role-playing preferences in cross-cultural settings, we set out to explore how these factors may be correlated. The first step is to match the two data sources according to the name list with identifiers. Given that participants were assigned with an identifier, they could remain anonymous while doing the thinking style measurement. During the interview, the coding process was likewise organized to ensure anonymity. The second step is to delete incomplete or irrelevant data. After completing these first two steps, data from 54 participants were determined valid. The last step was to calculate the correlation of thinking styles and roles. The results are shown in Table 6.

From the evidence, American students gained higher scores in hierarchical thinking styles (shown in Figure 1 and Table 4) and preferred to play the role of starter (shown in Figure 2 and Table 5). These findings are supported by correlation analysis that indicated that hierarchical thinking style has significant positive correlation with the starter ($p < 0.01; r = 0.363$). Accordingly, individuals who have obvious preference for hierarchical thinking style could focus on different things with different priorities in terms of goals (Sternberg, 1990), and thus are suitable to play the starter.

Table 6. Correlation coefficients of thinking styles and roles

<table>
<thead>
<tr>
<th></th>
<th>Starter</th>
<th>Supporter</th>
<th>Arguer</th>
<th>Questioner</th>
<th>Challenger</th>
<th>Timer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal</td>
<td>0.059</td>
<td>0.106</td>
<td>-0.156</td>
<td>-0.325*</td>
<td>-0.275*</td>
<td>0.247</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>0.363**</td>
<td>-0.149</td>
<td>-0.170</td>
<td>-0.379</td>
<td>-0.230</td>
<td>0.392**</td>
</tr>
<tr>
<td>Monarchical</td>
<td>0.008</td>
<td>0.292</td>
<td>0.228</td>
<td>0.06</td>
<td>0.101</td>
<td>0.004</td>
</tr>
<tr>
<td>Anarchic</td>
<td>0.045</td>
<td>0.212</td>
<td>-0.074</td>
<td>-0.113</td>
<td>-0.392**</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>0.750</td>
<td>0.132</td>
<td>0.604</td>
<td>0.425</td>
<td>0.004</td>
<td>0.476</td>
</tr>
<tr>
<td></td>
<td>0.012</td>
<td>0.062</td>
<td>-0.249</td>
<td>-0.355**</td>
<td>-0.246</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>0.935</td>
<td>0.660</td>
<td>0.075</td>
<td>0.010</td>
<td>0.078</td>
<td>0.589</td>
</tr>
</tbody>
</table>

Note. *$p < .05$, **$p < .01$.

American students obtained higher scores in anarchic thinking style and were not inclined to play questioner roles. The anarchic thinking style has a significant negative correlation with the questioner ($p < .01; r = -0.355$). Individuals who possess strong anarchic thinking style dislike rules, procedures, and guidelines (Sternberg, 1990). However, they enjoy working on tasks that allow flexibility as to what, where, when, and how one works rather than promoting a step-by-step process as planned (Zhang & Sternberg, 2000). The questioner is in charge of raising questions to facilitate discussion as planned. Therefore, the higher the anarchic thinking style, the more unlikely a student will play the role of questioner.
Individuals who have liberal thinking style usually enjoy engaging in tasks that involve novelty and ambiguity (Zhang & Sternberg, 2000). During CSCL, they must propose new ideas, new methods, or solutions. Both negative correlations were found between the liberal thinking style and the role of questioner \( (p < .05; r = -.325) \), and the liberal thinking style with the role of challenger \( (p < .05; r = -0.275) \). In this study, American students are clearly characterized as having liberal thinking style. Consequently, American students did not adopt the roles of questioners or challengers.

In cross-cultural CSCL, emerging roles vary widely with thinking styles. From this study, we have seen that Chinese students preferred roles, such as executive, global, conservative, and oligarchic, while also playing supporter, arguer, questioner, and challenger. American students meanwhile, tended to legislative, judicial, liberal, hierarchical, and anarchic, and were more likely to play starter and timer. The correlation analysis explored the relationship between emerging roles and thinking styles, that is

- a hierarchical thinking style has significant positive correlation with the starter and with the timer;
- an anarchic thinking style has significant negative correlation with the questioner; and,
- a liberal thinking style has negative correlation with the questioner and with the challenger.

**Discussion**

Understanding cross-cultural factors in joint activity is necessary in the age of globalization, in which demands are high for complex social engagement in various learning and performance settings. Knowledge of and sensitivity to cultural differences among different parties, especially in terms of learning dispositions, could help instructors facilitate collaborating participants with deliberately designed strategies and resources. These practices could promote better learning outcomes.

In this exploratory study, students from two cultural backgrounds were investigated based on their culturally related characteristics, the emerging roles they assumed in the collaborative learning process, and the relationships between the cultural factors and roles they played. Although not conclusive, these results reveal that cultural differences exist. In particular, the roles that students assume autonomously are culturally distinct and culturally related thinking styles that are significantly related with the adoption of certain collaborative roles.

Confirming the findings of Zhong (2010), the results reveal that cultural differences could be understood in terms of differences in thinking styles, and that these differences could affect the collaborative process. The nature of the interaction of thinking styles and role adoption has revealed differences of cultural features exhibited with different thinking styles. For example, Chinese students are more likely to work on multiple activities in the service of multiple objectives but are less likely to enjoy setting priorities. Meanwhile, American students are far more judicial, liberal, hierarchical, and anarchic. These qualities suggest that American students are more willing to engage in innovative tasks, evaluate work of others, and are likely supportive.

Consistent with this finding, Chinese students tend toward adopting roles of arguer, questioner and challenger, which are consistent with the thinking style of the oligarchic. By contrast, American students actively assume roles of supporter, starter, and timer, and are characterized as outgoing, enthusiastic, and friendly to group members, which are consistent with judicial, liberal, and hierarchical thinking skills.

The correlation analysis between the roles that students assumed and their preferred thinking styles further confirm that these cultural characteristics and associated roles in collaboration represent significant relationships. In particular, this notion is shown in hierarchical thinking style and roles of starter and timer, the anarchic thinking style and role of questioner, and the liberal thinking style and roles of questioner and challenger.

In similar cross-cultural CSCL contexts, learning efficiency might be improved by assigning scripted roles with the teacher assigning roles that align with cultural dispositions. However, when efficacy is not a primary concern, utilizing cultural differences of students in a CSCL project could be considered from the perspective of alternate learning purposes. For instance, when experience of multi-cultural factors is included explicitly within the learning purposes, teachers may assign roles for students that might contradict their dispositions.

This study was necessarily narrow and exploratory in its scope. Cultural features were approximated with thinking styles. Hence, this study makes no claim that the findings could be generalized. Nonetheless, this limitation has produced useful findings that indicate activities for future research. Because cultural features are complicated with multiple dimensions (Hofstede, 1980; 2011), more in-depth studies are required that might tease out which cultural dimensions might affect collaborative engagement other than thinking styles.
Conclusion

Cultural difference is expressed in many ways. Under the context of CSCL, this study reveals that thinking styles (Hofstede, 1980; 2011) plausibly approximate these differences. However, because of the limitations of the cohort size, further research is required to validate the findings from this study.

The findings suggest that in cross-cultural CSCL (or dPBL) contexts where certain individuals set the rules, make proposals, and deal with unexpected problems, assigning scripted roles according to cultural backgrounds may have value in promoting efficacy of the collaboration and in informing the design of cross-cultural studies.

References


Philosophy of Technology Assumptions in Educational Technology Leadership

Mark David Webster
Office of the Vice President of Technology, J. Sargeant Reynolds Community College, Richmond, VA, USA // mwebster@reynolds.edu

(Submitted September 6, 2015; Revised February 24, 2016; Accepted April 15, 2016)

ABSTRACT

A qualitative study using grounded theory methods was conducted to (a) examine what philosophy of technology assumptions are present in the thinking of K-12 technology leaders, (b) investigate how the assumptions may influence technology decision making, and (c) explore whether technological determinist assumptions are present. Subjects involved technology directors and instructional technology specialists from school districts, and data collection involved interviews and a written questionnaire. Three broad philosophy of technology views were widely held by participants, including an instrumental view of technology, technological optimism, and a technological determinist perspective that sees technological change as inevitable. Technology leaders were guided by two main approaches to technology decision making in cognitive dissonance with each other, represented by the categories Educational goals and curriculum should drive technology, and Keep up with Technology (or be left behind). The researcher concluded that as leaders deal with their perceived experience of the inevitability of technological change, and their concern for preparing students for a technological future, the core category Keep up with technology (or be left behind) is given the greater weight in technology decision making. A risk is that this can on occasion mean a quickness to adopt technology for the sake of technology, without aligning the technology implementation with educational goals.

Keywords
Educational technology leadership, Philosophy of technology, Instrumental view of technology, Technological determinism, Technological imperative, Technological optimism

Introduction

Philosophy of technology is a branch of philosophy that involves examining the underlying assumptions of how technologies impact and transform human society in ways that are philosophically relevant (Kaplan, 2009). Does philosophy matter? The author humorously recalls that Collier (1994) argued that proceeding without examining assumptions does not mean an absence of philosophy, but more likely, bad philosophy. Does philosophy of technology matter for educational technology leadership? Scholars have argued that philosophy of technology assumptions, especially technological determinism, may influence the thinking or discourse of educators concerning technology (Bennett & Maton, 2010; Cukier, Ngwenyama, Bauer, & Middleton, 2009; Fisher, 2006; Jones & Czerniewicz, 2010; Jones & Healing, 2010; Kanuka, 2008; Kritt & Winegar, 2010; Leonardi, 2008; Oliver, 2011; Selwyn, 2010; Strobel & Tillberg-Webb, 2009), and affect policy (Clegg, Hudson, & Steel, 2003; Cukier et al., 2009; Fisher, 2006; Wyatt, 2008). The pace of technological change presents challenges for contemplative leadership because there may be little time for considered judgment, with leaders responding in a reflexive rather than a reflective way to new information and change (Canole, 2007; Selwyn, 2010).

An important theme in the literature has involved scholarship discussing how technological determinism, a view pervasive in the popular mindset, may influence educational technology professional practice, shape interactions among stakeholders, and sometimes influence decisions. What is technological determinism? Technological determinism is the philosophical perspective that assumes that technology causes inevitable change in society (Leonardi, 2008; Leonardi, 2009), exerting a control over human society with technology considered in some way to be an autonomous force operating outside of social control (Feenberg, 2010; Hofmann, 2006; Leonardi, 2009). Technological determinism typically considers technology as a dominant force for social change, although there are different accounts of technological determinism.

Fisher (2006) examined discourse and rhetoric about educational transformation and observed a tendency for some discourse to be framed in technological deterministic language that ascribed to technology the power to inevitably cause positive change in schools. Fisher concluded that such technological deterministic assumptions are a problem because by ascribing autonomous change to technology, rather than to educators, the perspective shortchanges the hard work that educators must undertake to improve and transform education. If educators assume a commercial technology is inevitable, they tend to focus on how schools should adapt to technology, rather than shape the technology to meet curriculum requirements, and teachers’ and students’ needs (Jones &
Czerniewicz, 2010). In their critique of the claim that young people are digital natives who are naturally more able to use technology, with inevitable technology causing changes in students, Jones and Healing (2010) asserted that the digital native argument proceeds from a simplistic view of causality influenced by technological determinism. Kanuka (2008) argued that by examining their philosophy of technology assumptions, thoughtful practitioners with responsibilities for educational technology would be better able to make purposeful and informed decisions in selecting the right technologies for the right reasons.

Although educational technology scholars have emphasized the importance of critically examining philosophy of technology assumptions such as technological determinism, the researcher found that empirical studies within K-12 education were lacking. There was a gap in the literature concerning how technological determinist assumptions may influence the actual practice of educational technology leadership. However, philosophy of technology assumptions have been the focus of research outside of K-12 education to examine the influence of assumptions such as technological determinism on technology management and leadership. Technological determinist assumptions were found to influence the thinking of company managers, including their perceived agency in shaping technological change, affect discourse with colleagues and stakeholders, and influence the decisions that managers make on behalf of their organizations (Grant, Hall, Wailes, & Wright, 2006; Jackson & Philip, 2010; Leonardi, 2008; Leonardi & Jackson, 2004).

Seminal research by Leonardi and Jackson (2004) involving technology managers in the private sector found that discourse characterized by technological determinism, and the inevitability of technology, served as a powerful narrative to justify the actions of organizational leaders. Leonardi (2008) concluded that a “discourse of inevitability” created an ideological orientation toward technological change (p. 975). When technology managers employed technological determinist discourse, the tendency was to make the indeterminate state of things appear to be determined because of the perceived inevitability of technological change (Leonardi, 2008). Rhetoric characterized by assumptions of technological determinism can be a powerful discursive strategy for advancing interests and marginalizing dissenting opinions (Cukier et al., 2009; Leonardi, 2008; Leonardi & Jackson, 2004).

Oliver (2011) and Selwyn (2010) argued that research is necessary within education that critically questions technological determinist assumptions, and seeks to consider alternate ways of thinking about technology and learning that recognizes the agency of human actors, and the social factors involved with using technology in education. In responding to this research problem, the author pursued an empirical study guided by the work of Strobel and Tillberg-Webb (2009) who proposed a framework for educational technology that emphasizes a critical and humanistic approach to technology integration, and the importance of educators questioning their philosophical belief systems and values concerning technology. The starting point for the Strobel and Tillberg-Webb (2009) framework involves educators examining whether technological determinist assumptions influence thinking about educational technology.

The purpose of the study was to (a) examine what philosophical assumptions about technology are present in the thinking of K-12 technology leaders, (b) investigate how the assumptions may influence technology decision making, and (c) explore whether technological determinist assumptions are present.

Three research questions guided the qualitative study:
- What broad philosophy of technology assumptions are present in the thinking of K-12 technology directors and instructional technology specialists?
- How do philosophy of technology assumptions influence the decisions that leaders make about educational technology?
- What assumptions characterized by technological determinism may be present in leaders’ thinking or decision making?

The first research question was framed broadly so that the study would be open to any philosophical assumptions about technology. The second research question moved from examining thinking, to investigate how assumptions may influence decision making about technology. The third research question focused on examining any assumptions characterized by technological determinism.

**Research design and methods**

The research design aligned with Corbin and Strauss (2008) methods for analyzing data and developing grounded theory. Participants included school district technology directors, and instructional technology
specialists, and were selected through purposive and theoretical sampling. Data collection initially proceeded with purposive sampling to select 20 participants who work in educational technology leadership, with half of them school district technology directors and half instructional technology specialists. After the purposive sampling, the sampling shifted to theoretical sampling to develop the conceptual categories and emerging theory. Data collection continued until theoretical saturation had been reached, at which point 31 subjects had participated in the study, including 15 technology directors and 16 instructional technology specialists. Among the 31 total participants were 17 men and 14 women, from 19 school districts from different geographic areas throughout Virginia, including both city and county school districts.

In order to enhance the validity and reliability of the qualitative study, triangulation of data was pursued using interviews followed by a written questionnaire. The interviews used a semi-structured protocol, with most interviews conducted over the telephone. The written questionnaire with open-ended questions was completed by participants after the interview, and was emailed to them, completed, and then returned via email. Written questionnaires were returned by all 31 participants. Data collection resulted in a total of 31 interview transcripts and 31 written questionnaires, so that a total of 62 documents were analyzed.

The data analysis employed Corbin and Strauss (2008) grounded theory methods, based on the research guide Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. This article focuses on presenting the implications of the study for educational technology, while providing only a summary of the grounded theory methods used, as a full treatment of the research methods was previously published (Webster, 2016). The methodology employed many traditional grounded theory methods including constant comparative analysis, open and axial coding, use of the Corbin and Strauss coding paradigm, and theoretical saturation of categories. However, it should be noted that Corbin and Strauss methodology does not follow the classic grounded theory emphasis, in the Glaser tradition, of having no predetermined research problem. Corbin and Strauss (2008) methods for qualitative data analysis are accommodating of institutional requirements for research questions, literature review prior to research, and a theoretical framework. The author utilized the software application MAXQDA 11 to import transcripts, code conceptual categories, properties, and dimensions, to write memos while thinking critically about the data, and to refine conceptual theory.

Findings and implications for research question 1

Research Question 1 asked, “What broad philosophy of technology assumptions are present in the thinking of K-12 technology directors and instructional technology specialists?” The findings showed that three categories representing broad philosophy of technology views were prevalent among the educational technology leaders, including Technology is a tool, Technological change is inevitable, and Technological optimism. One category, Technology raises questions of human values, while not as prevalent as the other three, was associated with a majority of participants.

Participants often described how their philosophy of technology was characterized by viewing technology as a tool. For example, Technology Specialist 7 stated, “One philosophical belief of mine, shared by many of my educator acquaintances, is that technology is nothing more than a tool.” The 203 code instances for this category were distributed among 40 of the 62 documents, with coding instances from 27 of the 31 participants. The category Technology is a tool was interpreted as aligning with the instrumental view of technology, a philosophical view that considers technology as a tool, as means put to use by users for their purposeful ends (Berger, 2011; Feenberg, 1991; Heidegger, 2009). The implication is that this overarching philosophical perspective prevalent among the subjects is in alignment with what scholars and philosophers have argued is a popular and widely held philosophy of technology (Franssen, Lokhorst, & van de Poel, 2013; Jackson, 2010). The category Technology is a tool was found to be directly linked to the category Educational goals and curriculum should drive technology (explained below in the implications for Research Question 2).

A surprising finding of the study was that while the instrumental view of technology was prevalent, a philosophical position normally associated with the instrumental view, the neutrality thesis, was only associated with eight participants. The neutrality thesis considers technology as neutral with regard to values and assumes there are no inherent moral implications in using technology (Vermaas et al., 2011). On the other hand, the category Technology raises questions of human values was associated with a majority of the participants, 22 of 31. For example, Technology Director 10 stated, “Nowadays there’s a greater need to understand the ethical implications of using technology,” and gave the example of the need to honor ownership of digital property as an ethical consideration important for students and researchers. This subject asserted the opinion that free availability of information has desensitized some people to ownership of intellectual property.
The researcher interpreted the philosophical view represented by this category to be in alignment with Hofmann’s theory that technology is value laden, meaning it raises questions of values and has ethical consequences (Hofmann, 2006). Hofmann’s position does not mean differentiating between good technology and bad technology, rather the emphasis is that technology raises questions of human values, either through promoting particular values, or because the employment of technology has ethical consequences, whether intended or unintended (Hofmann, 2006). The data showed that the technology leaders described that technology, particularly in the context of K-12 education, raises questions of values as it relates to concerns such as Internet safety, personal privacy, equitable access to technology, copyright compliance, and environmental sustainability.

The surprising finding that the instrumental view of technology was a prevalent perspective, but the neutrality thesis normally associated with it was only associated with eight participants, might be interpreted as a possible case of cognitive dissonance between *technology is value neutral* and *Technology raises questions of human values*. Festinger’s theory of cognitive dissonance posits that persons experience discomfort or unease when they encounter new information that contradicts their previously held assumptions or beliefs (Sullivan, 2009). When a person considers two views that are inconsistent with each other, cognitive dissonance posits that the person weighs the importance of the cognitions, and may add or subtract from the alternatives to reduce the inconsistency (Harmon-Jones, 2009). Figure 1 displays the relationship between the contrasting philosophical perspectives *Technology is value neutral* and *Technology raises questions of human values*, and the greater weight given to the latter (Webster, 2013).

![Figure 1. Technology is value neutral vs. Technology raises questions of human values](image)

Discourse about the inevitability of technological change was a consistent refrain in the data. The philosophy of technology assumption *Technological change is inevitable* was prevalent and associated with 30 participants, with 149 coding instances distributed among 36 of the 62 documents. An example of a coding instance for this category is that Technology Specialist 2 stated, “Technological change is inevitable and we should not resist it. That is my philosophy! We’ve gone through more change because of technology than anything else in the last 150 years.” Within the broader category, properties or dimensions that had relatively high coding frequency included *BYOD is or seems inevitable*, *we should not resist technological change*, and *we should embrace technological change*.

Leonardi (2008) found from a study of technology managers that a “discourse of inevitability created an ideological orientation toward technological change” (p. 975). A similar phenomenon of an ideological orientation toward technological change is evident in this study of educational technology leaders. The philosophical view that technology causes inevitable change in society is an assumption associated with technological determinism (Leonardi, 2008; Leonardi, 2009). Discussion of research findings and implications pertaining to technological determinism are presented below in the section for Research Question 3.

The technology leaders participating in the study were generally optimistic about the potential for technology to improve education and the world, and they embraced its possibilities. Technological optimism was found to be a prevalent philosophy of technology, associated with 28 of 31 participants, with 137 coding instances distributed among 38 of the 62 documents. For example, Technology Director 15 stated, “A favorite saying of mine is that whatever the ill might be, technology will save the world!”

The Strobel and Tillberg-Webb (2009) humanizing framework for educational technology emphasizes the importance of critically analyzing whether assumptions about technology may correspond to viewpoints in the dichotomy of technological utopianism or dystopianism. Technological utopianism embraces the promise of technology, and is an optimistic position that presents technological innovation as something for the better (Kritt & Winegar, 2010; Strobel & Tillberg-Webb, 2009; Vermaas, Kroes, van de Poel, Franssen, & Houkes, 2011). The
contrasting perspective, technological dystopianism is a pessimistic position generally not open to technological innovation that resists technological change (Kritt & Winegar, 2010; Strobel & Tillberg-Webb, 2009; Vermaas et al., 2011). The category Technological optimism was interpreted by the researcher as aligning with the perspective of technological utopianism. The category Technological pessimism, interpreted to be aligned with technological dystopianism, involved 14 coding instances from 6 documents and 6 participants.

An important implication of Technological optimism is that it inclines technology leaders to be advocates and promoters for new applications of technology. A notable property of Technological optimism was technology advocacy. Technology Specialist 10 stated, “I try to stay optimistic and attempt to be proactive in encouraging my colleagues to use technologies.” Technological optimism also appears to be a strong factor in the leaders’ ideological orientation to technological change previously discussed. For example, Technology Specialist 9 expressed an optimistic feeling about technology in saying, “it’s a wonderful thing, and I have good feelings about technology.” In the same discussion, the technology leader described an orientation to technological change by sharing, “I think that technology is a wonderful thing that’s ever changing, it’s vital to have in education. It’s something that impacts our daily lives, I have really embraced it throughout my career.”

Findings and implications for research question 2

Research Question 2 asked, “How do philosophy of technology assumptions influence the decisions that leaders make about educational technology?” To obtain data pertinent to this research question, technology leaders were asked specific questions during the interview, and also in the written questionnaire, designed to connect philosophical thinking about technology to educational technology leadership or technology decision making.

Three categories related to technology decision making were found to be prevalent, Educational goals and curriculum drive technology, Keep up with technology (or be left behind), and Consider ethical factors associated with technology. Educational goals and curriculum should drive technology emerged as a widespread approach to technology decision making, involving 235 coding instances from 56 documents and 30 out of 31 participants. For example, Technology Director 8 stated, “There are no such things as technology initiatives! There are business of education projects designed to meet goals and objectives of student achievement through the use of technology.”

The approach to decision making represented by the category Educational goals and curriculum drive technology was found to be linked with the philosophy Technology is a tool, and an instrumental view of technology focused on considering technology as means put to use by users for purposeful ends (Berger, 2011; Feenberg, 1991; Heidegger, 2009). Such an instrumental approach to viewing technology education would accord with the views of scholars such as Jones and Czerniewicz (2010) who asserted that leaders should shape technology to suit educational needs and requirements. With the close link in the data between Educational goals and curriculum drive technology and the philosophy Technology is a tool, the implication is the instrumental view of technology leads to a corresponding approach to decision making that logically follows from the parent philosophy.

Participants also articulated the principle that technology should not be pursued for the sake of technology. Technology Director 4 explained, “I’m not a fan of technology for its own sake, and as a decision maker I like to see reasons for implementing technology.” A related property was curriculum should drive technology, rather than technology drive curriculum. For example, Technology Director 12 asserted,

You should not have the technology as the center of attention, but rather what it’s able to do, and allows your students to do. So I think the instructional component needs to be the focus, and how can we use the technology to make that happen.

An implication of the findings is that this perspective appears to be in alignment with what might be described as the textbook definition of appropriate technology integration, whereby curriculum drives technology integration (Shelly, Cashman, Gunter, & Gunter, 2008). Figure 2 depicts the instrumental view of technology as an overarching philosophy of technology, and its relationship to the decision making practice Educational goals and curriculum should drive technology (Webster, 2013).

Although Educational goals and curriculum should drive technology was a prevalent approach to technology decision making, the core category and central phenomenon that emerged from the study was that technology leaders approach technology leadership through a practice of Keep up with technology (or be left behind). This
category emerged early in data analysis, continued to grow in explanatory relevance, and reappeared in-vivo in transcript narratives in various forms, often expressed in those words, or variations of them. For example, Technology Specialist 11 stated, “As fast as technology changes, in education we should keep up with it or be one step ahead, but we’re usually one step behind.” Technology Director 3 shared, “Technology is always changing and you must change with it or you will be left behind.”

Figure 2. Instrumental view of technology and its relationship to the decision making practice Educational goals and curriculum should drive technology.
The core category was closely linked in the data with the category Technological change is inevitable, and a perceived imperative for schools to keep up with technological change. For example, Technology Specialist 8 stated, “My philosophy is that technology is imperative for today’s schools and for other sectors, it’s growing in leaps and bounds. Not all technology is good, but it’s an unstoppable force, and it has to be used and harnessed properly.” The core category was interpreted as a manifestation of a philosophy of technology perspective called the technological imperative, described by some scholars as associated with technological determinism (Chandler, 1995; Cukier et al., 2009; Hofmann, 2006). The technological imperative involves rhetoric and underlying assumptions that technology has a controlling influence (Hofmann, 2006) that is inevitable and unstoppable (Chandler, 1995; Cukier et al., 2009; Leonardi, 2008) and creates an imperative to keep up with technological developments (Strobel & Tillberg-Webb, 2009). The evidence indicates that the philosophy Keep up with technology (or be left behind) influences leaders by essentially functioning as an ideological orientation toward technological change. As mentioned earlier, this ideological orientation appears similar to what Leonardi (2008) found in a study of technology managers in the private sector.

Pressure to keep up with technology was the most frequent property of Keep up with technology (or be left behind), with 107 coding instances from 38 documents, and 26 of the 31 participants. Technology leaders often described how they felt pressured to keep up with technological change and the challenges that this presented for their work as educational technology leaders. The data suggests this pressure has a strong emotional interaction in the experience of many technology leaders. For example, Technology Specialist 9, who worked in a smaller school district with limited resources, stated about technology, “You have to always keep up with it. If you let any time go buy you’ll get behind and be lost.” The subject observed that a dilemma of keeping pace with technological change with fewer resources meant, “things move more slowly, and lamented that “sometimes when we are ready to proceed forward, we’re already behind, but we do a good job trying to keep up.”

Another property associated with the core category was resistance to technological change, which involved 76 coding instances, from 29 documents and 23 participants. A majority of participants described how the phenomenon of resistance to technological change was present in their school district. Technology Director 12 asserted education “tends to be very, very resistant to change, you’ve got a lot of folks who are not as comfortable with technology as the kids are, and that’s a big divide there, and that holds us back.” Like the core category itself, this property associated with it emerged at the beginning of data collection and analysis. The first participant interviewed, Technology Specialist 1, stated, “There can be a resistance to change in schools,” and asserted “resistance to technological change will keep us behind.”

The core category had two main properties in conflict with each other, pressure to keep up with technology, and the resistance to technological change they encounter in schools. The study’s results, showing pressure to implement technology occurring alongside resistance, appears to lend support to Rogers’ model of technological innovation, which theorizes that technological innovation proceeds alongside resistance from users in varying degrees (Friesen, 2008). The theme of resistance to technological change emerging from this study is similar to the resistance to change theme found by Cukier et al. (2009) in their qualitative study that used critical analysis to examine media discourse surrounding a technology initiative at a Canadian university.

Figure 3 depicts Keep up with technology (or be left behind) and its properties as a technological imperative following from the philosophy of technology assumption Technological change is inevitable (Webster, 2013). It shows the tension between pressure to keep up with technology and resistance to technological change. An implication of the findings is that the property Resistance to technological change had consequences, including teachers rarely or reluctantly integrate technology, and teacher resistance to technology disadvantages students. Pressure to keep up with technology was found to lead to an emphasis on preparing students for a technological future, but can lead to technology taking precedence over other values or norms (examples shown in figure). Weighted priority is placed by educational technology leaders on pressure to keep up with technology, as they struggle with resistance to technological change in their organizations.

Another finding pertaining to how philosophy of technology assumptions influence the decisions that leaders make about educational technology involves the category Consider ethical factors associated with technology, which involved coding instances from 29 participants. Technology leaders described how ethical considerations pertaining to technology were taken into account in making decisions about technology. For example, Technology Specialist 10 stated, “I often make technology decisions based on ethics, especially with teaching at the elementary level.” Technology Director 7 stated, “I believe as leaders we have an overall responsibility to consider ethics in our decision making.” Technology Specialist 9 explained, “Being in education, there are many things that we consider in terms of acceptable use policy, understanding age appropriate use of technology, parental permission, and ensuring safeguards.” Of the specific ethical considerations taken into account by the
technology leaders, one property emerged from a majority of the participants, consider Internet safety for students, with 53 coding instances from 25 documents, and 20 participants. An implication of the findings is that the philosophical perspective associated with this category is similar to the approach to technology and values taken in the national framework for K-12 educational technology leadership (Consortium for School Networking, 2011). The technology leaders in the study were concerned with ethical considerations pertaining to technology leadership similar to those issues defined in the Consortium for School Networking’s leadership framework.

Findings and implications for research question 3

Research Question 3 asked, “What assumptions characterized by technological determinism may be present in leaders’ thinking or decision making?” Most notably, the prevalent philosophical view Technological change is inevitable is an assumption related to technological determinism (Leonardi, 2008; Leonardi, 2009). The core

![Diagram](image-url)
category, *Keep up with technology (or be left behind)*, was linked in the data with viewing technological change as inevitable, and a perceived imperative within schools to keep up with this technological change. For example, Technology Specialist 8 stated:

My philosophy is that technology is imperative for today’s schools and for other sectors, it’s growing in leaps and bounds. Not all technology is good, but it’s an unstoppable force, and it has to be used and harnessed properly.

As explained previously, the core category was interpreted to be a manifestation of the technological imperative, a philosophical assumption associated with technological determinism (Chandler, 1995, Cukier et al., 2009; Hofmann, 2006). The technological imperative assumes that once technological development is inevitably underway, users should learn to cope with it (Chandler, 1995) because they cannot help but use technology (Leonardi, 2008), and should keep pace with technological change (Strobel & Tillberg-Webb, 2009).

An implication of the findings is that the phenomenon of the core category, *Keep up with technology (or be left behind)*, as a manifestation of the technological imperative, is similar to what was found in a qualitative study conducted by Cukier et al. (2009). Using content analysis techniques, Cukier et al. (2009) examined media discourse surrounding an instructional technology initiative at a university in Canada, and found that a technological determinist viewpoint was present in both academic and non-academic literature. In that study, rhetoric of the technological imperative was a dominant metaphor surrounding the technology initiative (Cukier et al., 2009). Discourse characterized by the technological imperative and the inevitability of technology can be employed to persuade others, with the rhetoric creating an ideological orientation in a culture toward technological change (Cukier et al., 2009; Leonardi, 2008). The findings from this study suggest that rhetoric surrounding *Keep up with technology (or be left behind)*, creates a “discourse of inevitability” (Leonardi, 2008, p. 975) in schools, and contributes to promoting an ideological orientation to technology within K-12 culture. For example, in describing discourse with colleagues in their schools, Technology Specialist 8 stated, “Everyone has to figure out how to make it work to aligning technology with clear educational goals, and essentially adopting technology for its own sake. A technological determinist viewpoint is present in both academic and non-academic literature.”

Conclusions and recommendations

The research findings led to a somewhat surprising substantive theory. As previously discussed, two dominant philosophical approaches were found to be important for educational technology leadership and decision making, *Educational goals and curriculum should drive technology*, and *Keep up with technology (or be left behind)*. The two approaches to technology decision making follow from their respective parent philosophy. *Educational goals and curriculum should drive* technology proceeds out of the broader philosophy *Technology is a tool* (instrumental view of technology), and from the perspective that technology is not an end in itself, but rather is a tool, a means to achieve educational goals and ends. *Keep up with technology (or be left behind)* proceeds out of a technological determinist perspective focused on the view *Technological change is inevitable*. Both of these parent philosophies are situated within *Technological optimism*. The irony is that these two philosophical approaches to technology decision making, were often held by the same technology leader, at the same time. The researcher concludes that as educational technology leaders respond to their perceived experience of the inevitability of technological change, and their concern for preparing students for a technological future, *Keep up with technology (or be left behind)* emerges as the primary of concern of leaders, and is given the greater weight in technology decision making. Figure 4 depicts the substantive theory that emerged from this study (Webster, 2013).

In shedding light on the philosophical tension between the two dominant approaches to technology decision making, it’s important to point out that under the instrumental view of technology, technology is employed as a means to an end, not an end itself, and not for its own sake. In contrast, when viewed from the perspective of inevitable technology, participants described how we should not resist, but should embrace technological change, and there can be a quickness to adopt technology for the sake of technology. A conclusion is that the pressures leaders experience to keep up with technology can result in procuring and implementing technology without aligning technology with clear educational goals, and essentially adopting technology for its own sake. A consequence of *Keep up with technology (or be left behind)* was well expressed by Technology Specialist 1 who observed that when technology is adopted for its own sake, “Everyone has to figure out how to make it work to support division and school needs. In most cases, this leads to resistance from teachers and usually dooms the technology to failure.” Technology Specialist 15 observed, “It seems like in the ever-evolving technology world,
folks are fast to jump on the bandwagon for the latest and greatest gadget or piece of software without first considering its instructional impact.”

![Diagram](image)

**Figure 4.** In competition and philosophical tension with Educational goals and curriculum should drive technology, Keep up with technology (or be left behind) is given the greater weight in technology decision making.

The budgetary dilemma for technology leaders feeling pressured to keep up with technology was well summarized by Technology Specialist 15, “You’re never going to be able to keep up with technological change in education because of our numbers of students and the cost.” Schrum and Levin (2009) advised that educators should wait for research demonstrating educational benefits before pursuing extensive technology investments. Besides the monetary consequences of extensive technology investments, there is the risk of quick technology adoptions being pursued without waiting for educational research to make an informed decision. Technology Specialist 2 wrote:

I would like to say that it is solid research that influences me, but I don’t need research to see students get excited using response systems, iPads, and interactive whiteboards. The game has changed and research cannot keep up with the changing tide, and I don’t want my students left behind.

As educators, our concern for preparing students for a future where technology competence is likely to be an advantage for them is certainly to be commended. Making a reasoned judgment to be an early adopter of innovative technology in schools may prove fruitful and beneficial to students. However, it can be argued that weighing *Keep up with technology (or be left behind)* too heavily over *Educational goals and curriculum should drive technology* may further disconnect technology from a focus on educational goals, and create further resistance in our organizations to technological change. With many educational concerns competing for limited fiscal resources, we have to be diligent in our efforts to ensure that our technology investments benefit administrative and academic functions in schools.

As discussed earlier, although educational technology scholars have emphasized the importance of critically examining philosophy of technology assumptions such as technological determinism, the researcher found that empirical studies within K-12 education were lacking. It is believed that this qualitative study broke new ground by filling a gap in the literature concerning how philosophy of technology assumptions, including technological determinist assumptions, may influence the actual practice of educational technology leadership. The Strobel and Tillberg-Webb (2009) framework, which guided this study, while based on good arguments and scholarship, was not grounded in empirical research. Therefore, this study’s findings and conclusions lend support to their theoretical framework. This study provides empirical evidence to support the concern of Strobel and Tillberg-
Webb that educators should critique their own beliefs and assumptions about technology, question deeper the connection between technology and human values, and engage in critical dialogue with other educators and students concerning such beliefs. This study sought to inform professional practice by contributing to what Kanuka (2008) called philosophy in practice pertaining to technology. We can conclude that philosophy of technology assumptions matter, that assumptions do shape leaders’ approaches to technology decision making, and by questioning philosophy of technology assumptions, technology leaders are better able to make purposeful and informed decisions.

While technology directors and instructional technology specialists provide leadership for educational technology, a limitation of this study is that these positions are not the only K-12 educators who provide leadership for integrating technology into instruction. It is recommended that similar qualitative studies be conducted involving other groups of educators, including principals or other central office administrators. It is also recommended that similar qualitative studies be conducted involving educational technology leaders or other groups of educators in other states or nations, both in primary and secondary education, and also in higher education. Because this qualitative study broke new ground, a limitation is that as future qualitative studies are conducted in the substantive area, it may become necessary to refine the substantive theory to accommodate new data. It is also recommended that data from this or other qualitative studies be used to help develop and validate a quantitative instrument to measure philosophy of technology assumptions, for use in quantitative research.

References


Mobile Learning in Pre-Kindergarten: Using Student Feedback to Inform Practice

Jennifer L. Reeves1*, Glenda A. Gunter2 and Candace Lacey3

1Nova Southeastern University, North Miami Beach, FL, USA // 2University of Central Florida, Orlando, FL, USA // 3Southeast Research and Evaluation Associates, Hollywood, FL, USA // jennreev@nova.edu // Glenda.Gunter@ucf.edu // chlacey@bellsouth.net

*Corresponding author

(Submitted September 8, 2015; Revised February 2, 2016; Accepted April 7, 2016)

ABSTRACT

There is a trend to use mobile devices in K-12 classrooms and create 1:1 learning experiences. Current research has focused on creating student collaborative efforts and increasing engagement when learning using the iPad, as well as the user-friendly characteristics and the tremendous number of apps available. There continues to be a need for empirical evidence supporting the effectiveness of mobile devices on student achievement. The purpose of this study was to determine how integrating mobile devices into a Pre-Kindergarten curriculum using informal feedback from students affects students’ academic achievement. The study employed a two-group, quasi-experimental design consisting of 28 students from two Pre-K classrooms. The experimental group utilized iPads with guided instruction, coupled with informal feedback from students, to target emergent literacy and early math skills; the control group did not have access to iPads. All students were given the Florida VPK Assessment at the beginning and end of the study. Results of the ANCOVA revealed significantly higher Phonological Awareness and Mathematics measures for the iPad group, suggesting that integrating mobile learning in content-specific areas using informal student feedback effectively increases early childhood education students’ academic achievement. Best practices for integrating mobile learning to enhance student engagement are discussed.

Keywords
Mobile learning, Tablets, Early childhood education, Formative assessment, Informal feedback

Introduction

In the pervasiveness of today’s digital society, mobile devices are changing the face of education. Although the Apple iPad was originally envisioned as a single user device, school districts and universities across the country are purchasing tablets and other mobile devices in massive quantities for innovative use in classrooms (Lawrence, 2012; Price, Jewitt, & Crescenzi, 2015; Tate, 2012). There is a great deal of anecdotal evidence (i.e., success stories) on the effectiveness of mobile devices in the classroom (Clarke & Abbott, 2013; Lenovo Education, 2012b; Mahaley, 2013; Morelock, 2011) and best practices for getting started (Victoria Department of Education and Early Childhood Development, 2011; Lenovo Education, 2012a; Mahaley, 2013); and in the last few years there are more researchers examining the effectiveness of mobile devices on student achievement (Chou, Block, & Jesness, 2012; Kucirkova, Messer, Sheehy, & Fernández Panadero, 2014; Ross, Morrison, & Lowther, 2010). The prevailing scholarly literature on mobile learning reflects positive correlations between student engagement and effective use of technology (Chen, Lambert, & Guirdy, 2010; Falloon & Khoo, 2014; Laird & Kuh, 2005), perceived engagement and perceived learning while using iPads (Diemer, Fernandez, & Streepey, 2012), and comfort levels and prolonged use of the iPad (Diemer et al., 2012).

While many of these studies were conducted in higher education institutions, there continues to be the need to focus on research in K-12 schools, especially in the early childhood education setting. Many of the previous studies have focused on user-friendly interface, mobility, touchscreen, and connectivity to a world of resources and apps (Price et al., 2015). Falloon (2013) conducted a study in New Zealand with 18 five-year-olds to investigate if using iPads with students could support learning. Falloon (2013) concluded all apps used for instruction should be evaluated for design and content prior to implementation in the classroom. Falloon (2013) also found that the disruption of an innovation and novelty of the devices could be used to motivate students and create engagement for innovative learning opportunities. Finally, Falloon (2013) alluded to the importance of incorporating formative and corrective feedback with students in the learning process.

Other researchers have focused on mobile devices to create student collaborative efforts and increase engagement when learning using the iPad. Kucirkova et al. (2014) examined the effects of a story-making app using 41 four- and five-year-old students from Madrid, Spain. Supported by previous research, Kucirkova et al. (2014) found that specific app features and content could influence the degree to which children’s engagement in the learning task was of “educational value” (p. 176). This research supported the notion that the iPad is a game
changers in education and that we need to find more ways to harness it as a powerful teaching and learning tool. Hwang and Chang (2011) conducted a study in Taiwan using formative assessment to determine if students’ attitudes and achievement could be affected when using mobile devices. They found that utilizing a mobile learning approach could improve students’ attitudes and achievement.

Due to the many promising learning opportunities, there continues to be a need for empirical evidence supporting the effectiveness of these devices on specific areas of student achievement. Therefore, the purpose of this study was to determine how integrating mobile devices into a Pre-Kindergarten curriculum using informal feedback from students affects Pre-K students’ academic achievement (i.e., Print Knowledge, Phonological Awareness, Mathematics, Oral Language/Vocabulary) at a rural, public charter school in the southeast.

Mobile devices, which include tablets, smartphones, PDAs, e-readers, and MP3 players, can be used to customize student learning and address education inequality when integrated into the classroom by innovatively informing policy and practice. Schrum and Glassett (2009) advocated that information and communication technologies could play a crucial role in empowering students to demonstrate authentic, meaningful learning. According to Chou et al. (2012), “with sound pedagogy and implementation, one-to-one learning has the potential to transform the classroom into a true learner-centered environment in which communication, collaboration, and creative problem solving flourish to create student-driven learning” (p. 13). Mobile technologies also are being used to address education inequality through various policies and practices throughout the United States. Many districts are attempting to close the achievement gap by giving students with limited social and economic resources access to mobile devices for use at school and home. Miami-Dade County Public Schools (M-DCPS, 2013), the fourth largest school district in the US, which consists of students from 160 countries, speaking 56 different languages, and 74% of the students being eligible for free and reduced lunch, approved a policy which ensured that every student had a digital device by 2015 (Smiley, 2013). According to Sung, Chang, and Liu (2016) teaching and learning with mobile devices is innovative and unique: they encourage anywhere, anytime learning; reach underserved children; develop 21st century social interactions; adapt to various learning environments; and enable a personalized learning experience.

Although little research exists on using informal feedback from elementary students to inform practice, Stiggins (2006) makes the distinction between assessment of learning (i.e., summative assessment) and assessment for learning (i.e., formative assessment). Stiggins argues that ongoing, informal feedback from students (i.e., formative assessment) improves students’ understanding of fundamental skills. Similarly, Andrade, Hefferen, and Palma (2014) found utilizing informal feedback among art teachers and students enhanced engagement and the quality of students’ work. In other words, informal feedback can be used to promote learning. The present study sought to determine how the integration of mobile devices in a Pre-K classroom, with the use of formative assessment, could enhance current instructional practices and create new opportunities for learning.

Theoretical framework

This study is grounded in the Theory of Mobile Learning (TML), which is based on constructivism and posits that learning is continuous and extends outside the classroom (Sharples, Arnedillo-Sánchez, Milrad, & Vavoula, 2009). Mobile devices enable 21st century students to construct their knowledge anywhere and anytime. According to Zimmerman and Howard (2013), “mobile devices can situate and connect learners by supporting authentic, context-specific, immediate learning” (p. 2). Thus, when students learn foundational literacy skills and basic mathematical concepts using mobile devices at an early age their learning takes on an entirely new perspective. Learning is no longer confined to the classroom; integrating mobile technology enables educators to customize student learning by creating authentic and appealing learning activities to engage students anytime and anywhere (Hess & Gunter, 2013).

Utilizing the Florida VPK assessment areas (i.e., Print Knowledge, Phonological Awareness, Mathematics, Oral Language/Vocabulary), which are aligned with the Florida Early Learning and Developmental Standards for Four-Year Olds (Florida Department of Education [FLDOE], 2013a), the following research questions guided this study: (a) How does mobile learning using informal feedback from students to inform practice increase Pre-K students’ Print Knowledge? (b) How does mobile learning using informal feedback from students to inform practice increase Pre-K students’ Phonological Awareness? (c) How does mobile learning using informal feedback from students to inform practice increase Pre-K students’ Mathematics skills? (d) How does mobile learning using informal feedback from students to inform practice increase Pre-K students’ Oral Language/Vocabulary skills?
The corresponding directional alternate hypotheses were (a) integration of mobile technology using informal feedback from students to inform practice will increase Pre-K students’ Print Knowledge, (b) integration of mobile technology using informal feedback from students to inform practice will increase Pre-K students’ Phonological Awareness, (c) integration of mobile technology using informal feedback from students to inform practice will increase Pre-K students’ Mathematics skills, and (d) integration of mobile technology using informal feedback from students to inform practice will increase Pre-K students’ Oral Language/Vocabulary skills.

Methods

Participants

Participants for the study consisted of 28 Voluntary Prekindergarten (VPK) students at a rural, public charter school in Florida. All children who reside in Florida and turn 4 by September 1st are eligible for Florida’s free VPK education program (Early Learning Coalition of Miami-Dade/Monroe, 2014). Participants were students from two Pre-K classrooms at the rural, public charter school. All participants were between the ages of 4 and 5; the majority of participants were Caucasian and from low to middle income families. Participant selection was based on cluster sampling using intact classrooms. The first classroom (n = 20) utilized iPads to practice their emergent literacy and early math skills. The second classroom (n = 8) did not have access to iPads in the classroom to practice skills.

Instrument

The Florida Center for Reading Research, in collaboration with the FLDOE, developed the Florida VPK Assessment in early 2012 to provide teachers with valid and reliable feedback regarding a child’s progress (FLDOE, 2013b). Based on the Florida Early Learning and Developmental Standards for Four-Year-Olds (FLDOE, Office of Early Learning, 2011), the VPK Assessment was created for teachers to guide instructional decisions in the Pre-K classroom based on progress monitoring. Although there are three assessment periods (Assessment Period 1 [AP1] given in September, Assessment Period 2 [AP2] given in January, and Assessment Period 3 [AP3] given at the end of May), only AP1 and AP3 are required. Therefore, for this study only AP1 and AP3 were analyzed.

The VPK Assessment includes progress monitoring measures in four areas aligned with the Florida Early Learning and Developmental Standards for Four-Year-Olds: (a) Print Knowledge (12 assessment items and 2 practice items), which assesses the child’s ability to recognize letters, name letters, and identify letter sounds; (b) Phonological Awareness (14 assessment items and 2 practice items), which assesses the child’s ability to combine smaller sounds or syllables to form a word, blend compound words, and recognize a word when part of it is taken away; (c) Mathematics (13 assessment items), which assesses counting skills, numerical associations, and math reasoning skills; and (d) Oral Language/Vocabulary (22 items for AP1 and 23 items for AP2 and AP3), which assesses a child’s expressive and receptive language, and targets the child’s knowledge of the parts of speech (FLDOE, 2013a). The content for the Florida VPK Assessment was developed by identifying the best predictors of later reading success (i.e., print awareness, phonological awareness, and oral language/vocabulary skills) and the best predictors of mathematics success (i.e., number sense; FLDOE, 2013a).

Data collection

The present study utilized a quasi-experimental design. Participants consisted of 28 students at a rural, public charter school in Florida. Participants were students from two PreK classrooms. All students were given the VPK Assessment (AP1) in September 2012. Upon IRB approval, beginning in October 2012, the experimental group (n = 20) practiced emergent literacy and early math skills utilizing iPads 2 days a week for approximately 15 minutes of guided instruction by one of the researchers, with the assistance of the VPK teacher. Over the course of the school year the following types of Applications (apps) were used with the experimental group: (a) Multi skill apps (e.g., phonics, counting, and matching), (b) Counting apps, (c) Early Math skill apps, (d) ABC apps, (e) Phonics apps, (f) Rhyming apps, (g) Matching Letters/Spelling apps, (h) Sight Word apps, (i) Tracing apps, (j) Science apps, and (k) Sorting/speed skill apps. With the unique diversity of mobile learning, the researchers and VPK teacher could choose basic Pre-K apps based on student skill levels to enhance each child’s individual abilities and needs. Informal feedback was collected from the students for each mobile app used. If students were
highly engaged in the app (i.e., on task for the entire time, eager to “play” on the iPads the next day, self-selecting the app when given a choice), the app was appropriate (i.e., no rude sounds for incorrect answers to disrupt the classroom, students were appropriately challenged), and it appeared to increase their knowledge on the subject area (i.e., students thought through the problem and chose the correct answer rather than simply going through the motions with little thought), then the app was approved and saved on the classroom iPads. Apps that did not meet the aforementioned criteria were deleted from the classroom iPads.

The experimental group continued to receive guided instruction using mobile devices until the end of the school year (i.e., for a total of 7 months). The control group (n = 8) did not have access to iPads in the classroom to practice skills; these students were taught with the traditional Pre-K curriculum. At the end of May, students in both the control and experimental groups were administered the Florida VPK Assessment (AP3).

**Data analysis**

Quinn and Keogh (2002) recommend using Type III sum of squares (SS) for unbalanced designs (i.e., those with unequal sample sizes). In SPSS, Type III sum of squares is the default method for unbalanced models with no empty cells (i.e., SPSS automatically accounts for the differences in the sum of squares by using Type III sum of squares so long as there is at least one observation per cell) and is equivalent to Yates’ weighted-squares-of-means technique (IBM, 2011). Although unequal sample sizes can be problematic, instead of randomly removing observations to equate samples, which could reduce power, Quinn and Keogh (2002) suggest ensuring the assumptions of the statistical analyses are met to minimize confounding issues. Although ANOVA is considered robust to moderate departures from the homogeneity assumption, when the sample sizes are very different, the departure needs to be minimal (Sweet & Grace-Martin, 2012).

In order to determine whether integrating mobile technology into the Pre-K curriculum using informal feedback from students’ enhanced emergent literacy and early math skills, an Analysis of Covariance (ANCOVA) was conducted using AP1 (i.e., the pretest) as the covariate. Prior to the analysis, viewing a normal Quantile-Quantile (Q-Q) probability plot, which plots the observed values to the expected normal values, assessed the assumption of normality. When the normality assumption is met, the observed values and the expected normal values approach a straight line. The Q-Q plots for all Florida VPK assessment measures (i.e., Print Knowledge, Phonological Awareness, Mathematics, and Oral Language) showed little deviation from the line, suggesting the normality assumption was met for all measures. In addition, the Levene’s test for equality of error variances was met for all variables and Tukey’s Least Significant Difference (LSD) was used to control the familywise error rate.

**Results**

To ensure the experimental and control groups were similar at the outset, independent samples t tests were conducted for each of the AP1 measures. Results revealed no significant differences between the experimental and control groups in any of the AP1 measures (see Table 1) ensuring that both the experimental group and the control group were similar at the beginning of the year.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>M</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Knowledge</td>
<td>iPad</td>
<td>8.30</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>6.86</td>
<td>1.42</td>
<td>.30</td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>iPad</td>
<td>6.65</td>
<td>.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.71</td>
<td>1.64</td>
<td>.52</td>
</tr>
<tr>
<td>Math</td>
<td>iPad</td>
<td>9.50</td>
<td>.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>7.89</td>
<td>2.25</td>
<td>.39</td>
</tr>
<tr>
<td>Oral Language</td>
<td>iPad</td>
<td>16.60</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>15.71</td>
<td>1.77</td>
<td>.61</td>
</tr>
</tbody>
</table>

*Note. VPK = Voluntary Prekindergarten; AP1 = Assessment Period 1.*

Results of the ANCOVA revealed a significant testing effect, Wilk’s $\lambda = .26, F (4, 22) = 15.60, p < .001, \eta^2 = .74, \beta = 1.00$; see Table 2). Not surprisingly, both experimental and control groups performed significantly better on all AP3 measures compared to AP1 measures.
Results of the ANCOVA also revealed a significant testing by class effect, Wilks’ $\lambda = .52$, $F\left(4, 22\right) = 5.00$, $p = .005$, $\eta^2 = .48$, $\beta = .91$ (see Table 3). Students in the experimental group had significantly higher Phonological Awareness and Mathematics measures during AP3 compared to the control group; consequently the null hypotheses for Phonological Awareness and Mathematics were rejected. Even though the experimental group outperformed the control group during AP3 for Print Knowledge and Oral Language, the results were not significant. Consequently, the researchers failed to reject the null hypotheses for Print Knowledge and Oral Language.

Table 2. Mean Florida VPK Assessment scores

<table>
<thead>
<tr>
<th>Measure</th>
<th>Testing</th>
<th>$M$</th>
<th>SE</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Knowledge</td>
<td>AP1</td>
<td>7.60</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AP3</td>
<td>11.05</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>AP1</td>
<td>6.18</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AP3</td>
<td>9.18</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>AP1</td>
<td>8.68</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AP3</td>
<td>12.9</td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td>Oral Language</td>
<td>AP1</td>
<td>16.16</td>
<td>.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AP3</td>
<td>18.95</td>
<td>.59</td>
<td></td>
</tr>
</tbody>
</table>

Note. AP1 = Assessment Period 1 (i.e., pretest); AP3 = Assessment Period 3 (i.e., posttest). *$p < .05$.

Table 3. Mean Florida VPK Assessment scores during AP3

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>$M$</th>
<th>SE</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Knowledge</td>
<td>iPad</td>
<td>11.25</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>10.86</td>
<td>.64</td>
<td>.17</td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>iPad</td>
<td>11.50</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>6.86</td>
<td>1.13</td>
<td>.02*</td>
</tr>
<tr>
<td>Math</td>
<td>iPad</td>
<td>14.70</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>11.14</td>
<td>.97</td>
<td>.04*</td>
</tr>
<tr>
<td>Oral Language</td>
<td>iPad</td>
<td>20.05</td>
<td>.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>17.87</td>
<td>1.03</td>
<td>.12</td>
</tr>
</tbody>
</table>

Note. *$p < .05$.

Discussion

According to research conducted by the FLDOE (2013a), students can benefit from targeted instruction in the four skill areas assessed in the study (i.e., Print Knowledge, Phonological Awareness, Mathematics, and Oral Language) during the Pre-K year. The purpose of this study was to target emergent literacy and early math skills for a group of Pre-K students using mobile technology to enhance instruction, use informal feedback from students to guide app selection, and determine how interactive mobile learning affected academic achievement, as measured by the Florida VPK Assessment. Results of the study suggested that mobile learning using informal feedback from students to guide instruction significantly increased students’ Phonological Awareness and Mathematics skills compared to a control group that did not receive targeted instruction using mobile technology. Since many of the iPad apps used with the experimental group specifically focused on phonics (i.e., phonological awareness) and counting, sequencing, or early addition (i.e., mathematics), this result was anticipated. These findings are consistent with many of the informal success stories on mobile learning (e.g., Clark & Abbott, 2013; Lenovo Education, 2012b; Mahaley, 2013; Morelock, 2011), and add to the literature by providing empirical evidence on the power of innovative teaching practices, student-centered formative assessments, and interactive instructional strategies.

Mobile learning using informal feedback from students had no significant effect on students’ Print Knowledge, which was inconsistent with the study’s hypotheses. Since print knowledge is a basic Pre-K skill that all students must master, this result was not surprising. In addition, at the beginning of the study the teacher of the control group classroom asked if she could use the Starfall Education Website, which specifically focuses on print knowledge, on the classroom desktop computer with her students. Since this was one of her standard instructional strategies, the researchers allowed it since they did not want to hinder the development of the control group. Therefore, focused attention on print knowledge by the control group, regardless of medium, could have minimized the differences between the groups. It is also important to note that Starfall was one of the iPad apps used with the experimental group that was used regularly based on positive, informal feedback from students. Given both the experimental and control groups improved their print knowledge significantly, it
appears print knowledge-specific iPad apps and the Starfall Education Website are both effective at enhancing Pre-K students’ print knowledge. Finally, mobile learning using informal feedback from students had no significant effect on students’ Oral Language/Vocabulary skills. Since the research did not specifically use apps that focused on expressive and receptive language, or targeted knowledge of parts of speech, this result was also not surprising. Consequently, this confirms that only the areas targeted by the researchers were affected by the use and integration of iPads in the classroom. Given the small number of participants in the study, future research should replicate the study using larger sample sizes that are equivalent, and guided instruction focused on oral language skills to determine if mobile learning can increase achievement in this area as well.

Recommendations

Based on the findings of this study, the researchers suggest the following recommendations for future studies and educators who plan to implement mobile devices as a teaching strategy. Prior to allowing students to use the mobile devices, teachers need to plan extra time for evaluating and setting up iPads or mobile devices for an integrated lesson. Young students are quite savvy and will quickly determine how to delete apps. Restrictions on the iPad should be enabled to make sure only age-appropriate content can be accessed. The researchers recommend that educators in early childhood should turn off the ability to delete apps. According to Lacey, Gunter, and Reeves (2014), choosing appropriate apps is the first step to engaging the digital learner. Although there are many free apps out there, many of them offer in-app purchases that are distracting for young students. The researchers recommend turning off the ability to make in-app purchases in restrictions area, since the ads are very tempting to young minds. All apps should always be evaluated and tested prior to using them in the classroom. The researchers found that some of the apps that were used in this study provided inappropriate feedback and noises for incorrect answers. Other apps do not give the proper remedial assistance. Through the informal feedback process it was quickly discovered that these types of apps should not be included on the classroom iPads. When selecting educational apps the researchers suggest seeking other types of evaluation or sites such as APPtic, which is a directory of educational apps that have been tested extensively by Apple Distinguished Educators.

Once a variety of apps have been selected, the next step is to introduce them to students slowly. Starting with one app at a time, make sure to have several different levels available to appropriately differentiate instruction. When working with younger students, educators should plan to introduce apps for a short period of time (about 10-15 minutes) due to young student’s shorter attention spans. Students’ informal feedback can provide valuable insight in the evaluation process of the apps for instructional value. The informal feedback can be gathered overtly or covertly, depending on the age and cognitive abilities of the students. To evaluate whether the app will increase knowledge on the subject area educators must ensure that students are being challenged appropriately based on their level and the app is content-specific matched to the curriculum standards. Teachers should remove any apps from the classroom devices that do not meet the aforementioned criteria. The researchers recommend creating a list of approved apps for parents and posting on the classroom Website. Engaging parents in the process can improve achievement for all students, and for schools with BYOD initiatives, it ensures all students have the classroom-approved apps on their own devices. For mobile learning to be successful educators must provide opportunities for students to learn anytime and anywhere. Educators should consider other uses and recourses available with mobile learning, such as virtual field trips; situational learning with problem solving; interactive gamified learning; and student created projects, like digital stories and screen casting.

This study provides empirical evidence to support the integration of mobile technology in content-specific areas combined with the use of informal student feedback increases early childhood education students’ academic achievement, as well as offers suggestions on how to integrate mobile learning into the classroom. With today’s digital generation having access to knowledge at the tip of their fingers, anytime and anywhere, it is imperative that teachers integrate mobile learning into their teaching best practices. Additionally, in order to increase student engagement and achievement, early childhood educators should use informal feedback from students to inform classroom practice. With school districts across the country purchasing mobile devices in bundles in an attempt to close the achievement gap and engage at-risk students, educational researchers must continue to provide empirical evidence to illustrate the power of mobile learning.

References


Exploring Learners' Sequential Behavioral Patterns, Flow Experience, and Learning Performance in an Anti-Phishing Educational Game

Jerry Chih-Yuan Sun1*, Cian-Yu Kuo1, Huei-Tse Hou2 and Yu-Yan Lin1

1Institute of Education, National Chiao Tung University, Taiwan // 2Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taiwan // csun@ntu.edu.tw // metilda_kuo@kingston.com.tw // hthou@mail.ntust.edu.tw // aoisora.nagi@gmail.com

*Corresponding author

(Submitted September 10, 2015; Revised December 15, 2015; Accepted January 26, 2016)

ABSTRACT

The purposes of this study were to provide a game-based anti-phishing lesson to 110 elementary school students in Taiwan, explore their learning behavioral patterns, and investigate the effects of the flow states on their learning behavioral patterns and learning achievement. The study recorded behaviour logs, and applied a pre- and post-test on phishing knowledge and a flow state measurement to analyze the learning process. The study used lag sequential analysis to infer the students’ behavioural patterns. The results showed that the learning materials used in this study can enable learners’ flow experience, whereby they can acquire anti-phishing knowledge through trial and error via a repeated “learning with gaming” behavioral pattern. We recommend that future educators and researchers on this topic appropriately increase the level of difficulty of the games used, and design learning materials with flexible difficulty based on learners’ flow states.

Keywords

Game-based learning, Anti-phishing, Behavioral patterns, Flow experience, Sequential analysis

Introduction

As of 2014, there were 2.4 billion Internet users globally (Internet World Stats, 2014). However, studies show that Internet users are often unaware of the dangers of the phishing attacks they are exposed to (Arachchilage & Love, 2013). As of 2013, Internet phishing through email scams had decreased by 16% from 2012 (Fossi et al., 2013); however, Internet scams using social sites had increased by 3% (Anti-Phishing Working Group, 2013). Social networking websites are gradually replacing emails as the new primary means for Internet scams (Fossi et al., 2013). Victims of Internet phishing include people of all age groups, among which young people constitute the most vulnerable group (Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2010). Statistics from Livingstone and Haddon (2009) showed that 60% of all social site users were between 9–16 years old, while more recent data from 2013 further showed that the number of young people who were social sites users increased by 29% in one year and was continuing to increase (Harper, 2014). This heavy use of social sites accentuates young people’s vulnerability to phishing through these sites.

Tools for detecting phishing may not provide effective defenses (Kumaraguru et al., 2007a). Therefore, some academics deem anti-phishing education, in which users learn correct Internet phishing concepts and defense methods and are equipped with basic defense abilities, to be the most effective means of combating phishing (Kumaraguru et al., 2007b; Kumaraguru et al., 2010; Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010; Sheng et al., 2007) as well as fostering good general Internet safety, operating habits, and awareness (Wirth, Rifon, LaRose, & Lewis, 2007). Defense against Internet phishing is a problem-solving process that requires judgment and reasoning abilities (Kumaraguru et al., 2007a). The advantage of game-based learning materials is the construction of life-like situations of phishing and game-based learning scenarios; therefore, game-based learning is an effective teaching method to help learners enhance their anti-phishing abilities (Kiili, 2005; Sheng et al., 2007). The game-based learning method affects learners’ flow experience and problem solving abilities, and can promote effective learning behaviors; for example, simulation games help learners enter a flow state during the computational problem solving activities (Liu, Cheng, & Huang, 2011) and facilitate the flow experience when playing games as a group in an interactive social game environment (Inal & Çağıltay, 2007). There have been previous studies focusing on the use of anti-phishing games in teaching among university students (Yang, Tseng, Lee, Weng, & Chen, 2012), but there has been no study exploring the effective learning behavior of children or teenagers through anti-phishing educational games. Therefore, the present study, using young people as the research subjects and an anti-phishing game as the teaching method, sought to understand whether the flow experience in games can contribute to effective learning behaviors.
Mayer (1992) proposed that it is relatively difficult for teachers to understand learners’ metacognitive processes such as learning strategies through summative assessment (e.g., learning achievement tests), and that it is therefore difficult to improve learners’ learning states or enhance their learning achievement using these methods. Thus, it may be more effective to help educators understand and improve learners’ learning states and strategies using quantified dynamic learning behavior, presented in the form of a time order (Liu et al., 2011), as well as applying sequential analysis to explore the correlation between learners’ behaviors and to infer overall behavioral patterns (Bakeman & Gottman, 1997). Learning behavior patterns are beneficial in terms of identifying the reason for unsatisfactory learning states and learning achievement, which can in turn help teachers provide learners with suitable guidance for better learning outcomes (Hou, 2012b). On this basis, this study makes use of a game-based learning scenario and sequential analysis to explore learning behaviors in order to enhance young people’s anti-phishing abilities on social sites. During online learning activities, if the learning state reaches the condition of flow experience, learners will have greater learning achievement (Choi, Kim, & Kim, 2007; Skadberg & Kimmel, 2004). Liu et al. (2011) analyzed learning behavioral patterns in game-based learning and discovered that learners tended to use behavioral strategies learned from mistakes. Based on the above findings, a stronger understanding of the correlation between flow states and learning behavioral patterns in game-based learning should help future educators guide the learning behaviors of students and help them reach a higher level of learning achievement. Therefore, the purpose of this study was to examine differences between learners in terms of their learning behavioral patterns and learning achievement given different flow experiences relating to game-based anti-phishing learning materials.

**Literature review**

**Young people and Internet phishing on social sites**

“Internet phishing” refers to the illegal act of attackers making use of fake emails, websites, or browsers to deceive victims and obtain their private information when they click on links or fill in personal information (Sheng et al., 2007). According to past studies, approximately 90% of all Internet users eventually click on links to phishing sites within eight hours of the time the fake emails are sent, and thus fall victim to these illegal acts (Kumaraguru et al., 2009; Kumaraguru et al., 2007a); furthermore, most users are relatively unconcerned about issues relating to Internet phishing (Fossi et al., 2013). Social sites provide Internet users with communication channels to develop interpersonal relationships, generally free of charge (Gao et al., 2010); at the same time, they also create complicated social connections which facilitate phishing victimization (Boshmaf, Muslukhov, Beznosov, & Ripeanu, 2011).

Although children and youths are a digitally native generation, they are at greater risk from Internet phishing than are adults due to their lack of experience and knowledge relating to Internet scams (Fire, Goldschmidt, & Elovici, 2014). Young people are most frequently attacked by Internet scams on social sites (Wirth et al., 2007). Quilliam, Rifon, and Larose (2006) highlighted the fact that over 60% of young people fail to regularly look for privacy policies, 58% do not clear their browser history, 37.7% fail to check if online forms are secure, and 50% do not set their browsers to reject cookies; some youths even authorize third-party software or websites to obtain their personal information without reading the user agreements on those sites (Wirth et al., 2007). Therefore, a comprehensive anti-phishing curriculum is needed to help young people build their knowledge of Internet phishing so as to equip them with the needed defensive abilities and prevent them from becoming victims (Wirth et al., 2007).

Most current studies related to anti-phishing education have used university students as their teaching subjects, and only a few have focused on conducting Internet information safety education for children. Jansson and von Solms (2013) designed an educational training method focused on phishing mail that allowed learners to acquire anti-phishing related knowledge from simulated phishing attacks by receiving emails. The study by Robila and Ragucci (2006) taught university students awareness of the content of phishing attack websites through classroom discussions. Yang et al. (2012) developed an Anti-Phishing Education Game to develop university students’ knowledge of anti-phishing techniques; their results showed that students showed significant progress in their ability to identify phishing web pages. Moreover, with children aged between 9 and 12 as subjects and by means of questionnaire and behavioral observation, Wishart, Oades, and Morris (2007) explored whether online role-play activities could help children receive and remember the key Internet safety message.

To sum up the above observations, young people are the primary users of social sites, but at the same time lack knowledge and skills relating to Internet safety, and the ability to defend themselves against social site phishing, and therefore run a very high risk of becoming victims. Among the current studies of anti-phishing or Internet
information safety education, there are few focusing on anti-phishing education among children and teenagers. Research findings from Wishart et al. (2007) only listed the important concepts of Internet safety that children learned from role-play activities, such as: ‘Don’t give out personal details’ and ‘Don’t trust what people say in chatrooms/on the Internet;’ they did not, however, investigate the learners’ behavioral process in the course of role-play. As such, this study targets youths with an intervention to enhance their Internet safety awareness through anti-phishing game-based instruction, and to equip them with the ability to identify and defend against Internet phishing, while also examining their behavior patterns during the learning process.

**Flow experience in game-based learning**

Csíkszentmihályi (2000) considered flow to be the optimal experience, whereby actors are completely absorbed in flow activities, with their consciousness focused within a narrow range and automatically filtering out perceptions and feelings that are not related to the activity at hand. In a flow state, actors have a sense of control over their environment and are focused on specific goals and clear feedback. In the three-channel model of flow (flow, anxiety, and boredom) proposed by Csíkszentmihályi (2000), “flow” is a feeling of balance between skill and challenge, and its force can drive actors into challenging activities at a higher level. Figure 1 is an illustration of the three-channel model. When there is a balance between the level of challenge of the learning activity and the actor’s skill, the actor will experience a flow state; when an actor of low skill faces high-difficulty challenges, he or she will enter into an “anxiety state”; and when their skill is higher than the difficulty of the challenge, they will not be fully focused on the activity and will enter into a “boredom state” (Csíkszentmihályi, 2000). There are different techniques for measuring flow. One is the use of a psychological scale to detect the aspects of perceived control, engagement, and enjoyment following the conclusion of an activity (Pearce, Ainley, & Howard, 2005; Webster, Trevino, & Ryan, 1993), as an overall measure of flow state. Another method is to monitor learners’ skill-challenge balance during the activity, as an indicator of the flow state process (Pearce et al., 2005). Massimini and Carli (1988, p.269) describe this method as “theoretically, the most meaningful reference point for the presence or absence of flow.” Fang, Zhang, and Chan (2013) argued that all flow elements (including preconditions, flow state, and its outcomes) should be measured so that one can understand whether learners actually achieve the flow state. However Inal and Çağiltay (2007) found that children failed to understand some questions in the flow scale, such as the sub-factors of autotelic experience, time distortion and playability. Therefore, when considering young people’s comprehension ability, and to minimize the level of intervention in the learning process, this study used the method of measuring skill-challenge balance to explore learners’ flow state. Furthermore, this study focused on the correlation between flow state and behavior outcome in the learning process, so other flow related factors (such as preconditions) were not included in the measurement.

![Figure 1. Three-channel model of the flow state (Csíkszentmihályi, 2000, p.45)](image)

Although flow is an optimal experience, Csíkszentmihályi (1975) found that it only existed among certain kinds of activities such as climbing, chess and basketball. Among these specific activities in which people might experience flow, playing a game is an activity that allows one to undergo an excellent flow experience (Csíkszentmihályi, 1975, p. 36). As a result, this study considered game-based learning a suitable environment for designing the activities to facilitate flow experience. In digital learning, a game-based curriculum is viewed as computer-assisted scenario teaching and an effective tool for enhancing learning problem solving abilities (Liu et al., 2011). Game-based learning seeks to induce learners to repeatedly practice the concepts learned and to focus on learning tasks (Hou, 2012b) while striking a balance between skills and challenges (Sweetser &
Wyeth, 2005). Games provide appropriate levels of scaffolding and difficulties, enhancing learners’ performance in problem-solving abilities (Hung, Kuo, Sun, & Yu, 2014), while the challenge in games helps learners achieve better flow experience (Hung, Sun, & Yu, 2015). In a game-based learning environment, flow experience is a suitable factor for exploring the learning state of learners (Kiili, 2005; Liu et al., 2011). Therefore, the present study suggests that in a game-based learning environment, flow theory helps to illuminate the learning process and effectively categorize learners, and hence uses the three-channel model of flow to examine the differences in behavioral patterns of flow experience and their correlation with learning achievement.

The relation of flow experience with learning behavioral patterns and learning achievement

Educational games can promote effective learning behaviors (İnal & Çağultzay, 2007). Hou and Li (2014) suggested that game-based learning stimulated learning motivation using game-based learning materials, thereby encouraging learners to voluntarily engage in the activity at hand and achieve flow experience, as well as enhancing their learning achievement by providing enjoyable experiences and challenging goals. If learners are fully focused on the activity while maintaining a desire to achieve the goal (İnal & Çağultzay, 2007), they will be more effective at learning (Hoffman & Novak, 1996; Hou & Li, 2014). Therefore, maintaining the flow state can be considered an effective way of promoting game-based learning achievement (Liu et al., 2011). Pearce et al. (2005) discovered a positive correlation between flow experience and learning achievement, as flow experience enabled learners to achieve higher levels of learning achievement. It is thus no surprise that measuring the flow experience of learners in an online learning scenario is indeed beneficial to understanding learning state (Choi et al., 2007; Skadberg & Kimmel, 2004). This study expects that in a game-based environment, learners who have achieved a flow state will focus on the learning activities and proactively face challenges, allowing better learning achievement.

On the other hand, game-based learning also provides the opportunity to explore suitable learning methods through learners’ experiences (Liu et al., 2011). Therefore, studying learning behavioral patterns and related processes will help us discover the reasons behind unsatisfactory learning achievement and identify means of improvement (Hou, 2012b). Derived from the basis of observation of the learning process, learning behavioral patterns summarize the learners’ behavioral characteristics (Hou, 2013). Liu et al. (2011) analyzed learning behavioral patterns in game-based learning and discovered that learners tended to use behavioral strategies learned from mistakes: Learners who were less capable of understanding the questions and considering solutions would lose motivation and more easily fall into a state of anxiety, in turn prolonging their learning duration and reducing their learning achievement. A study by Hou (2013) pointed out that in game-based learning, learners with more prior knowledge were more confident at interacting with others and discussing questions, and achieved greater enjoyment, whereas learners with less prior knowledge lacked the confidence to interact with others and were more prone to choose to learn on their own. Based on the above findings, a stronger understanding of the correlation between flow states and learning behavioral patterns in game-based learning should help future educators guide the learning behaviors of students based on their flow experience, and help them reach a higher level of learning achievement. Therefore, this study uses game-based learning materials, and categorizes the flow state of learners based on the three-channel model of flow in order to analyze differences in learners’ learning behavioral patterns and learning achievement by flow experience.

Research methods

Participants and instructional design

This study is a causal-comparative research study. The participants were students in grades 5 and 6, aged between 9 and 12, with an average age of 11.23. There were 110 students altogether, of whom 62 were boys and 48 were girls. The research data were collected in December 2014. In order to control the learning environment to ensure that the students would not be influenced by external factors, we conducted a two-hour lesson in a computer lab, and the entire process was recorded on video.

A pre-test on phishing knowledge about social sites was conducted before the start of the lesson to collect information on the learners’ prior knowledge. Following that, the learners were informed of the learning task and its goals. The task required the students to learn according to their own methods during the anti-phishing game-based lesson; however, they were required to successfully pass the game challenge at least once before they could complete the learning task. Learners started challenging the learning task once the lesson started; they were allowed to try different learning behaviors, such as reading the learning materials, referring to practical
examples, and consulting forums until they succeeded in their challenge. Learning behaviors during the lesson were logged and coded automatically through the computer system. The lesson ended once the challenges were completed successfully, at which point an active survey was conducted immediately in order to measure the learners’ flow experience (Novak, Hoffman, & Yung, 1998), as was a post-test on phishing knowledge about social sites, to collect information on after-lesson learning achievement. The experimental procedures and scenarios are shown in Figures 2 and 3.

**Figure 2. Experimental design**

**Figure 3. Game scenario of anti-phishing**
The anti-phishing game-based lesson used in this study required learners to learn independently using digital learning materials, with both the lesson and the test being conducted entirely on computers. The anti-phishing game-based learning materials were related to phishing on social sites; the main system functions covered can be classified into six types, namely user identification, the learning material, practical examples, games, forums, and leader boards. Learners were able to register and log into the system through the user identification function. Learners tried to pass the game challenges by answering the questions, and were then rewarded with treasure. The questions were designed based on the anti-phishing concepts and were presented randomly. Prior research showed that game-based learners tend to learn from their mistakes (Liu et al., 2011), so this study was designed to allow the learners to continually work on the game challenges. During the learning process, they acquired knowledge related to phishing attacks on social sites by reading the learning materials, experiencing the practical examples and games, and using the forum for discussion with peers; they were also able to view the learning achievement of their peers through the leader board.

**Instruments**

The data-gathering tools used by this study were the flow state measurement, a pre- and post-test on phishing knowledge about social sites, and learning behavioral coding. We used the study of Pearce et al. (2005) as a reference for the flow state measurement, which measured the learners’ flow state during the activities using the challenge and skill questions. Learners were asked to answer two questions based on their feelings during the learning experience: (1) *Do you feel that your skills were suitable for solving the tasks in this activity?* and (2) *How do you feel about the task’s level of challenge in this activity?* The measurement used a five-point Likert-type scale (*1* = highly unsuitable; *5* = highly suitable). The measurement is a type of technique to probe the flow state of learning process, similar to the research of Pearce et al. (2005). In this research, one question was used for each of the two dimensions of skill and challenge to probe students’ flow space. Nevertheless, due to limitations of the course duration and to prevent interference with learners’ discussions with their peers during the learning process, reading and teaching materials, challenge games and flow state measurements were only determined following the completion of a learning task (that is, after successfully getting through one level of the game) as an indicator of flow space location.

The pre- and post-test on phishing knowledge about social sites (for summative evaluation) used the same questions—ten questions consisting of a mix of multiple-choice and yes/no questions, in a different sequence in the pre- and post-tests. We referred to social-site-phishing-related events for the question propositions, which were further examined and repeated by experts in digital learning. A preliminary test was taken by twelve youths to ascertain the item discrimination and difficulty level of the test. Based on the Ebel and Frisbie (1986) criterion, we amended two multiple-choice questions, as their difficulty levels were too high, and one giveaway multiple choice question, whose difficulty level was too high and whose item discrimination was insufficient. After the revision, the average difficulty level of the test was 0.56, and the discrimination power of D value was 0.54, which achieved excellent quality. Cronbach’s α was .64, acting as the minimum acceptable level of reliability for preliminary research (Nunnally, 1967) and reflecting learning achievement derived from these game-based anti-phishing learning materials related to social sites (Nunnally & Bernstein, 1994). The final test questions on phishing in social network sites are listed in Appendix 1.

Concerning the validity of the learning achievement test, three experts in digital learning and behavioral patterns examined the validity of the test questions so as to ensure expert validity. Expert A’s research focuses on coding learning behavior processes. He has investigated various models of learning behavior through quantitative content analysis and sequential analysis. Expert B’s research interests are in integrating information technology into teaching, designing digital teaching materials and action learning. Expert C’s research field includes teaching technology, interactive learning, online learning and course design, and designing multimedia teaching. The experts in digital learning and behavioral patterns reviewed the study and defined four behavioral dimensions (Reading, Interaction with peers, Game, and Test) before conducting the experiment. The students’ learning behaviors were then automatically recorded by the system in the anti-phishing curriculum (Hou, 2012a; 2013; Tsai, Yu, & Hsiao, 2007). Detailed definitions of the codes are provided in Appendix 2.

**Statistical analyses**

Upon completion of the data collection, the learners were first grouped based on information from the flow state measurement: learners with equal skill and challenges were grouped into the flow state; if skill was greater than challenge, the learners were grouped into the boredom state; if skill was lower than challenge, they were grouped
into the anxiety state. The data were analyzed by SPSS 17.0 and the Mepa 4.9 statistical software following file conversion. Frequency distribution was used for descriptive analysis, while one-way analysis of variance (ANOVA) was used to examine whether different flow experiences had any basis in differences in prior knowledge in the pre-test on phishing knowledge about social sites. Subsequently, a paired-samples t-test was performed to examine the average scores of different flow experiences in the post-test on phishing knowledge about social sites. In addition, we used sequential analysis to calculate a behavioral frequency conversion matrix for different flow experience states and converted the behavioral code information of the various groups into the z-score, followed by the export of the frequency transition tables and adjusted residuals tables. The z-score in the table was then used to determine the significance between the behavioral sequential relations (z > 1.96 indicated that the sequential relation reached the significance level of p < .05) (Bakeman & Gottman, 1997). Lastly, the results were organized into a learning behavioral patterns chart.

Research results

Descriptive statistics

Grouping the learners according to the flow state measurement showed that there were 60 learners in the Flow Group (32 boys and 28 girls, with an average age of 11.32), 32 learners in the Anxiety Group (19 boys and 13 girls, with an average age of 11.25), and 18 learners in the Boredom Group (11 boys and 7 girls, with an average age of 11.11). In terms of gaming experience, the respective numbers of learners with one year, two years, three years and more than three years of gaming experience were 40, 8, 14 and 48. In addition, 30 learners had used social networking sites previously, while the remaining 80 learners had not.

In this study, we collected a total of 2,362 behavioral codes (Flow Group: 1242, Anxiety Group: 721, Boredom Group: 399). In terms of dimensional difference, given the different numbers of learners in each group, after deducting the (game) success (Y) and failure (N) codes, the frequency distribution of their learning behaviors across various dimensions is shown in Table 1. In terms of the Reading (9%, 10%, and 11%) and Test (36%, 33%, and 38%) dimensions, the three groups were similar; as for peer Interaction, the behaviors of the Flow Group and the Anxiety Group were similar, while the level of interaction was notably lower in the Boredom Group. However, the Boredom Group’s interaction in the Game dimension was also notably higher than that for the other two groups.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Dimensions</th>
<th>Reading</th>
<th>Interaction with peers</th>
<th>Game</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow</td>
<td>Reading</td>
<td>107 (9%)</td>
<td>163 (14%)</td>
<td>492 (41%)</td>
<td>480 (36%)</td>
<td>1242</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Reading</td>
<td>64 (10%)</td>
<td>84 (13%)</td>
<td>289 (44%)</td>
<td>284 (33%)</td>
<td>721</td>
</tr>
<tr>
<td>Boredom</td>
<td>Reading</td>
<td>41 (11%)</td>
<td>9 (2%)</td>
<td>177 (49%)</td>
<td>172 (38%)</td>
<td>399</td>
</tr>
</tbody>
</table>

Learning behavioral patterns in different flow states

We first performed sequential analysis on the Flow Group, and the residual table was then adjusted based on the dimensions: significant sequences for $p < .05$ were Test→Test ($z = 26.18$), Reading→Reading ($z = 4.65$), Interaction→Reading ($z = 5.15$), Interaction→Interaction ($z = 26.60$) and Game→Game ($z = 21.07$). The learning behavioral patterns chart is shown in Figure 4. The Flow Group had three notably significant behavioral sequences: (1) the behavior of performing multiple Tests indicated that the learners had completed all the tests; (2) learners would keep the Game playing behavior until they succeeded; (3) learners continually interacted with their peers and engaged in Reading behavior when their interactions ended.

We then adjusted the residual table based on the dimensions of the Anxiety Group and discovered that the significant sequences for $p < .05$ were Test→Test ($z = 17.82$), Test→Reading ($z = 4.90$), Reading→Reading ($z = 3.12$), Interaction→Interaction ($z = 17.56$) and Game→Game ($z = 16.65$). This group’s learning behavioral patterns chart is shown in Figure 5. The learning behaviors of the Anxiety Group exhibited the following characteristics: (1) performing multiple Tests indicated that learners had completed all the tests; (2) after completing the pretest, learners started and continued to Reading; and (3) upon the completion of Test and Reading, learners would keep the Game playing behavior until they succeeded, while also Interacting repeatedly with peers.
Finally, we adjusted the residual table based on the dimensions of the Boredom Group, and discovered that the significant sequences for $p < .05$ were Test→Test ($z = 13.89$), Test→Reading ($z = 3.14$), Reading→Reading ($z = 7.24$), Interaction→Interaction ($z = 3.10$) and Game→Game ($z = 12.80$). This group’s learning behavioral patterns chart is shown in Figure 6. Their dimension behavior eventually returned to the Test dimension, indicating that the Boredom Group had completed all the tests in the study. The learning behaviors of the Boredom Group exhibited the following characteristics: (1) learners preferred repeated Reading behavior after the pretest; and (2) upon completion of the Test and Reading, learners would keep the Game playing behavior until they succeeded, while also Interacting repeatedly with peers. Thus, their behaviors were similar to the behaviors of the Anxiety Group.
Learning achievement in different flow states

We first used one-way ANOVA to examine whether different flow groups had significant differences in their average scores on the pre-test on phishing knowledge about social sites. Based on the application of Levene’s test, \( p = .21 \) was taken to mean that the assumption of homogeneity of variance was not rejected and that subsequent analysis could be performed. The ANOVA results showed that the pre-test scores of all the flow groups did not reach significance \( (F = 1.21, p = .30) \), indicating no significant difference in the prior knowledge of the three groups; therefore, the possibility of prior knowledge affecting learning achievement in the post-test could be ruled out, and subsequent analysis could proceed.

The results of the paired \( t \)-test are shown in Table 2. The overall pre- and post-test results on phishing knowledge about social sites reached significance \( (t = 3.02, p = .03) \). Among the groups, however, the Flow Group did not reach significance in the pre-test or the post-test \( (t = 1.28, p = .21) \); the Anxiety Group reached significance on both tests \( (t = 3.36, p < .01) \); and the Boredom Group did not reach significance on either test \( (t = 0.79, p = .44) \). In short, the anti-phishing game-based learning materials significantly improved the learning achievement of the Anxiety Group, while the other two group did not see any significant improvement in their learning achievement.

Table 2. Paired-samples \( t \)-test of various flow state groups’ anti-phishing learning achievements

<table>
<thead>
<tr>
<th>Groups</th>
<th>Pre-test</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>( t )</td>
<td>( p )</td>
<td>( d )</td>
</tr>
<tr>
<td>Flow</td>
<td>47.25</td>
<td>15.33</td>
<td>49.67</td>
<td>14.95</td>
<td>1.28</td>
<td>.21</td>
<td>.54</td>
</tr>
<tr>
<td>Anxiety</td>
<td>43.59</td>
<td>10.87</td>
<td>52.03</td>
<td>13.37</td>
<td>3.36</td>
<td>&lt; .01</td>
<td>.33</td>
</tr>
<tr>
<td>Boredom</td>
<td>49.72</td>
<td>15.58</td>
<td>52.50</td>
<td>12.86</td>
<td>0.79</td>
<td>.44</td>
<td>.47</td>
</tr>
<tr>
<td>Total</td>
<td>46.59</td>
<td>14.25</td>
<td>50.82</td>
<td>14.12</td>
<td>3.02</td>
<td>&lt;.03</td>
<td>.35</td>
</tr>
</tbody>
</table>

Note. The minimum anti-phishing achievement test score was 5, the maximum was 100, and the range was 95.

Discussion

Flow experience and learning behavioral patterns

This study’s results for learning behavioral patterns are similar to those of Liu et al. (2011), which proposed that learners tend to learn from their mistakes. After a failed game, the Flow Group might engage in the behavior of discussing the results with peers, followed by the Reading behavior, and would continue to repeat these behaviors until they succeeded in the learning task. In other words, the Flow Group not only engaged in independent learning but also sought assistance from peers (Hou & Li, 2014). However, the difference between our study and previous studies is that the learners in the Flow Group would verify the information by reading the materials again after discussing and interacting with peers. This may have been due to the different designs of the game-based learning environments. Liu et al. (2011) adopted a collaborative simulation game, while the current study incorporated a task-challenge style, helping learners promote a behavioral strategy of collaborating with others and reading/verifying materials. We infer that after discussing and interacting with peers, learners in the Flow Group would verify the outcomes of the discussion using the learning materials before they proceeded with the game, and that this is what made their learning more effective (İnal & Çağiltay, 2007). The Anxiety Group also sought help from peers after a failed game and kept repeating this behavior until they succeeded. We infer that learners in the Anxiety Group also tended to turn to reading materials and peers for solutions, but lacked the step of reading and verifying what they had learned from the discussions; instead, they simply continued playing the game (and thus had higher frequency of Game behavior compared to the Flow Group). As a result of this process of repeated trial and failure and the learning behaviors of overestimating the learners’ levels of understanding due to the lack of opportunity to internalize the knowledge (Csíkszentmihályi, 2000), this group experienced more anxiety than the others. Finally, the Boredom Group relied less on the reading materials and interaction with peers, tending instead to try the challenges repeatedly until they succeeded, again similar to the findings of Liu et al. (2011). Although Hou (2013) suggested that learners with less prior knowledge lack the confidence to interact with peers, the flow experience of the three groups in our study did not show significant differences in prior knowledge, ruling out this interpretation. Instead, it may be that this group’s level of prior knowledge and skill far exceeded the difficulty level of the tasks, leading to the feeling of boredom and making it difficult for them to focus on the activities (Csíkszentmihályi, 2000).
The influence of flow experience on anti-phishing game-based learning achievement

As for post-test learning achievement, only the Anxiety Group showed that the anti-phishing game-based learning materials were effective in improving anti-phishing knowledge in relation to social sites. This result is contrary to past results, which have proposed that the flow state should lead to positive learning achievement (Choi et al., 2007; Hou & Li, 2014; Pearce et al., 2005; Skadberg & Kimmel, 2004). With regard to the Flow Group, based on learning behavioral patterns, we infer that the reading behavior of this group was insignificant before the game, leaving the participants in this group with insufficient anti-phishing knowledge and ability to solve the game tasks. Instead, this group completed the games through peer assistance and re-reading, leading them not to perform well on the post-test in the absence of assistance. Although the Anxiety Group had read the learning materials before the game and substantially acquired the anti-phishing knowledge provided, this knowledge remained insufficient for them to complete the tasks without interacting with peers to strengthen their skills. Although the Anxiety Group lacked the behavior of verification after interaction and reading, leading them to require more time to complete the tasks, they were able to obtain relatively good learning achievement when allowed to read the materials comprehensively and engage in repeated practice. Finally, the materials read by the Boredom Group before the game were sufficient for them to complete the tasks, and they felt bored by the (putative) challenges of the lesson, lacked peer interaction, and were unable to fully focus on the activities. As a result, they showed no significant improvement in their learning achievement, a result congruent with the views proposed by Csikszentmihalyi (2000).

In terms of learning efficiency, the Flow Group enjoyed the games, and viewed their skills to be in equilibrium with the level of difficulty of the challenges (Kiili, 2005; Liu et al., 2011); they sought to accomplish the game tasks using suitable methods. Although the Flow Group’s incidence of Game behavior was slightly lower than that of the Anxiety and Boredom Groups, the combination of interaction and reading was an effective learning model. The results of our study are similar to those of past studies that have proposed that flow experience can help learners focus and maintain their motivation to reach their goals (İnal & Çağiltay, 2007), thereby leading to more effective learning (Hoffman & Novak, 1996; Hou & Li, 2014).

Conclusion, limitations, and recommendations for future study

The purpose of this study was to examine the differences in learning behavioral patterns and learning achievement among learners with different flow experiences. The results showed that learners can acquire anti-phishing knowledge through trial and error via a repeated gaming behavioral pattern. Learners in the Flow Group might engage in the behavior of discussing with peers and verifying the outcomes of the discussion; the Anxiety Group tended to turn to reading materials and peers for solutions, but lacked the step of reading and verifying what they had learned; the Boredom Group relied less on reading materials and interaction with peers, tending to try the challenges repeatedly until they succeeded. It is suggested that future educators and researchers on anti-phishing education appropriately increase the level of difficulty of games used and design learning materials with flexible difficulty based on learners’ flow states.

In terms of study limitations, flow was measured using only two questions in this study; thus, various factors potentially impacting flow experience were not considered. Due to insufficient flow scale matching, the age groups of our research subjects in this study, and the relatively short duration of the experiment, flow state in this study only focused on the skill-challenge balance measured on one occasion among students aged 9-12. To fully understand flow state, the learners’ flow precondition, state, and outcomes need to be measured (Fang et al., 2013). Pearce et al. (2005) also argued that the complex changes of flow state have to be measured continuously throughout the learning process through questions concerning skill and challenge. Therefore, we suggest that future studies may consider measuring flow experience before and during activities so as to better understand changes in learners’ flow states, in order to understand better how changes in flow affect learning behavior. Furthermore, although the materials of this study did lead to learning achievement, the learning content could still be improved to make the lesson more interesting and enhance learning motivation, for example by making the game tasks more challenging or more dynamic to stimulate the learning motivation and flow experience (Hung et al., 2014; Hung et al., 2015).

With regard to game-based learning, we suggest that future educators offer learning questions and guide learners to interact with their peers; further, they can encourage learners to verify information as well as to improve the involvement of learners experiencing boredom by getting them to help others, thereby helping them to form the flow experience from a new angle. Whether for the development of game-based learning materials or the delivery of related lessons and courses, learning content should allow learners to better enjoy reading the
materials and challenging themselves by playing the games, as well as to adjust the difficulty of games in a timely way according to flow state, so as to maintain learners’ flow experience and help them engage in activities with greater learning efficiency and effectiveness, which should in turn enhance the educational value of the game-based learning materials used.

Acknowledgements

This research was supported by the Ministry of Science and Technology in Taiwan through Grant numbers MOST 104-2511-S-009-008-MY3, MOST 103-2511-S-009-008-MY2, NSC 101-2511-S-009-010-MY3, and NSC 100-2511-S-009-012. The authors would like to thank the students who participated in this study and acknowledge the contributions of Shian-Shyong Tseng, Sunny S. J. Lin, and Chao-Hsiu Chen who supported this research study and provided valuable comments.

References


Appendix 1. Test questions on phishing in social network sites

<table>
<thead>
<tr>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True or False Questions</strong></td>
</tr>
<tr>
<td>T 1. After being added into a shopping group, it is correct to select “Report Group” in the top-right corner of the group’s page and then select the reason “This is junk email or fraud” and “This has caused inconvenience to me.”</td>
</tr>
<tr>
<td>T 2. It is correct to regularly select the “Setting” icon in the top-right corner of Facebook, execute the setting function, enter into the “Account Privacy” settings page, and select “Applications” to remove unused programs.</td>
</tr>
<tr>
<td>F 3. To report unlawful Facebook shopping groups, one may call the reporting line 0800-016587, the main line 02-2215071, or send a fax to 02-22142627; one may also write to No. 66, Anfeng Road, Xindian District, New Taipei City 23156, or send an email directly to <a href="mailto:liao@tipo.gov.tw">liao@tipo.gov.tw</a>.</td>
</tr>
<tr>
<td>F 4. When logging into Facebook, directly inputting the password (in the sequence of first character, second character…) could prevent Trojan horse programs from being imbedded into computers to record the users’ passwords.</td>
</tr>
<tr>
<td>F 5. It is best to ignore the situation when you discover that your account has been continuously inviting others to join a shopping group.</td>
</tr>
<tr>
<td>T 6. When using Google Chrome as your browser, it is correct to select “≡” → “Tools” → “Extensions” (you may type chrome://extensions/ directly) to regularly check, deactivate, and remove the “Chrome Service Pack” application.</td>
</tr>
<tr>
<td>T 7. When using Mozilla Firefox as your browser, it is correct to select “Tools” → “Extensions” (or use the keyboard shortcut Ctrl + Shift + A) and select “Extensions” to regularly check, deactivate, and remove the “Mozilla Service Pack” application.</td>
</tr>
<tr>
<td>T 8. Entering into the “Advertisement” setting page in account settings, and changing the settings for “Third-party Sites” and “Advertisement and Friends” to “Limit to my friends” ensures that when you “Like” any advertisements in the future, it will not become a channel for advertising among your friends.</td>
</tr>
<tr>
<td>F 9. When you see an instant notification from Facebook showing that “You have been tagged in a YouTube Video Link “Adobe Flash Click!!””, it means there is a required update and you should open the link without worrying.</td>
</tr>
<tr>
<td>T 10. Setting security codes on Facebook is one of the most effective methods to defend against phishing.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiple Choice Questions (MCQs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D 1. How do you activate a two-stage verification? Please arrange the steps in sequence: 1. Select setting 2. Account security 3. Account setting 4. Log in authorization 5. Check “require safety pin, allow me to access my account from an unknown browser”.</td>
</tr>
<tr>
<td>A 12345</td>
</tr>
<tr>
<td>B 13452</td>
</tr>
<tr>
<td>C 23145</td>
</tr>
<tr>
<td>D 13245</td>
</tr>
<tr>
<td>A 2. How do you recover your account within five minutes? Please arrange the steps in sequence: 1. Log into the website <a href="http://www.facebook.com/hacked">http://www.facebook.com/hacked</a> that was jointly developed by Facebook and the Crime Investigation Bureau (CIB) 2. Select “My account has been hacked” and enter the user’s account details, followed by an account detail confirmation 3. Enter the user’s existing or past passwords 4. Reset the password and increase passwords’ complexity 5. Set a new security question and recover account access.</td>
</tr>
<tr>
<td>A 12345</td>
</tr>
<tr>
<td>B 12435</td>
</tr>
<tr>
<td>C 14325</td>
</tr>
<tr>
<td>D 12543</td>
</tr>
<tr>
<td>C 3. Attackers use rewards or prizes to invite users to join dubious events, or fill in personal information to steal their information. Which of the following phishing tactics on social networking sites is the aforementioned tactic?</td>
</tr>
<tr>
<td>A Manual sharing of virus attack</td>
</tr>
<tr>
<td>B Dubious “Like” phishing attack</td>
</tr>
<tr>
<td>C Dubious reward phishing attack</td>
</tr>
</tbody>
</table>
4. When attackers tempt users into clicking on links to certain content or “Like”, a different action immediately takes place in the background. Which of the following is an example of this kind of phishing tactic?
   A. Dubious “Like” phishing attack
   B. Dubious reward phishing attack
   C. Manual sharing of virus attack
   D. Dubious browser phishing attack

5. Attackers create false news or interesting clips to tempt friends into clicking/selecting them, and collect their personal information. Which of the following is an example of this kind of phishing tactic?
   A. Dubious “Like” phishing attack
   B. Dubious reward phishing attack
   C. Manual sharing of virus attack
   D. Dubious browser phishing attack

6. Attackers usually use vouchers as rewards and invite users to their URLs to attach malicious “JavaScript” codes for activating the fraud mechanism. Which of the following is an example of this kind of phishing tactic?
   A. Dubious “Like” phishing attack
   B. Virus copy phishing attack
   C. Manual sharing of virus attack
   D. Dubious browser phishing attack

7. Attackers trick users into downloading malicious browsers, which then reasonably expand in user’s computers, stealing sensitive personal information during the process. Which of the following is an example of this kind of phishing tactic?
   A. Dubious “Like” phishing attack
   B. Dubious reward phishing attack
   C. Manual sharing of virus attack
   D. Dubious browser phishing attack

8. A frame is superimposed onto a dubious video player, once a user clicks on the video or “Like”, the fraud mechanism is activated immediately to collect the user’s information. Which of the following is an example of this kind of phishing tactic?
   A. Dubious “Like” phishing attack
   B. Dubious reward phishing attack
   C. Manual sharing of virus attack
   D. Dubious browser phishing attack

9. Which one of the following is “NOT” a method for defending against phishing tactics on social networking sites?
   A. Switch to non-sequential key-in of password during log in
   B. The best way is to reinstall the computer when you found yourself continuously inviting others to shopping groups
   C. Set up security codes
   D. Cancel log in notification
## Appendix 2. Game-based learning behavioral coding scheme

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Coding</th>
<th>Behavior</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading</strong></td>
<td>R</td>
<td>Reading course learning</td>
<td>Select the “Course learning—learn about social sites attack methods” page to read.</td>
</tr>
<tr>
<td></td>
<td>r</td>
<td>Reading course learning</td>
<td>Select the “Course learning—guard against social sites attack methods” page to read.</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Reading real-life practice</td>
<td>Select the “Real-life practice—learn about social sites attack methods” page to read.</td>
</tr>
<tr>
<td></td>
<td>e</td>
<td>Reading real-life practice</td>
<td>Select the “Real-life practice—guard against social sites attack methods” page to read.</td>
</tr>
<tr>
<td><strong>Interaction with peers</strong></td>
<td>C</td>
<td>View leaderboard</td>
<td>Select the “Leaderboard” page and view the challenge scores of other learners.</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>View message</td>
<td>Select the “Forum.”</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>Post question message</td>
<td>Select the “Forum” and add comments or post questions.</td>
</tr>
<tr>
<td><strong>Game</strong></td>
<td>G</td>
<td>Challenge games</td>
<td>Select the “Game” page.</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Successful challenge</td>
<td>Select the “Game” page and win the challenge.</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Failed challenge</td>
<td>Select the “Game” page but fail the challenge.</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td>S</td>
<td>Take the pre-test</td>
<td>Select the “Pre-test” page and take the test.</td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>Take the post-test</td>
<td>Select the “Post-test” page and take the test.</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>Take the flow test</td>
<td>Select the “Flow test” page and take the test.</td>
</tr>
</tbody>
</table>
Computerized Dynamic Adaptive Tests with Immediately Individualized Feedback for Primary School Mathematics Learning

Huey-Min Wu¹, Bor-Chen Kuo² and Su-Chen Wang³

¹Research Center for Testing and Assessment, National Academy for Educational Research, New Taipei City, Taiwan // ²Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taichung, Taiwan // ³Taiping District Sin-Guang Elementary School, Taichung, Taiwan

wuhm@mail.naer.edu.tw // kbc@mail.ntcu.edu.tw // mandy2539@gmail.com

*Corresponding author

(Submitted September 18, 2015; Revised December 2, 2015; Accepted January 22, 2016)

ABSTRACT

In this study, a computerized dynamic assessment test with both immediately individualized feedback and adaptively property was applied to Mathematics learning in primary school. For evaluating the effectiveness of the computerized dynamic adaptive test, the performances of three types of remedial instructions were compared by a pre-test/post-test nonequivalent group design. These remedial instructions are the adaptive dynamic assessment with individualized instruction, the individualized instruction without adaptive dynamic assessment, and the traditional classroom remedial instruction. One hundred and eighteen 5th grade elementary school students from six classes participated in the study and were assigned to three remedial instruction programs on a random basis. The results demonstrated that the computerized dynamic adaptive test with individualized prompts outperformed the other two instruction methods.

Keywords

Computerized adaptive test, Dynamic assessment, Graduated prompts, Knowledge structure, Mathematics

Introduction

With the use of computer and communication technology, researchers have developed various assessment systems in improving students’ learning achievement. The characteristic of learning or testing systems is to provide meaningful and personalized feedback that can accommodate a diversity of student personalization (Chen, 2009; Chen, 2011; Hwang, Panjaburee, Triampo, & Shih, 2013; Kao, Lin, & Chu, 2012; Luft, Gomes, Priori, & Takase, 2013; Moridis & Economides, 2012; Panjaburee, Hwang, Triampo & Shih, 2010). However, many computerized-based tests place greater emphasis on feedback on the final product of assessment tasks rather than the process of the learning. Moreover, all examinees received the same items and were required to complete all the items within the assessment procedure. These test systems do not capitalize on advantage of combining technology with testing, a computerized adaptive test (CAT). CATs offer examinees customized items suited to their aptitudes and cognitive status and shorten the length of the test (van der Linden & Pashley, 2010). Adaptive tests can provide equal measurement precision for most test-takers (Yen, Ho, Chen, Chou, & Chen, 2010; Yen, Ho, Liao, & Chen, 2012).

Wu, Kuo, and Yang (2012) have developed a knowledge structure-based computerized adaptive test for diagnosing students’ learning profiles. In the research of Wu et al. (2012), an adaptive test algorithm that can utilize fewer items than the original paper-based tests was proposed. The knowledge structure-based computerized adaptive test provides feedback regarding how well students performed on specific skills immediately after testing with fewer items (Wu et al., 2012); however, it is not sensitive to the process of students’ learning. Meaningful feedbacks that can facilitate learning effectiveness are timely feedback and process-oriented feedback (Gabelica, Van Den Bossche, Segers, & Gijselaers, 2012; Harks, Rakoczys, Hattie, Besser, & Klieme, 2014; Hattie & Timperley, 2007; Parr & Timperley, 2010; Wang, 2011).

Dynamic assessment is an interactive assessment that can provide students with both timely and process-oriented feedback (Campione & Brown, 1990; Haywood & Lidz, 2007). Dynamic assessment is grounded in Vygotsky’s (1978) sociocultural theory, especially the Zone of Proximal Development (ZPD) concept, which describes how a student can go from his actual development level to his potential development level through continuous communication and consultation in the course of interactions with teachers, peers, or parents (Poehner, 2008). The ZPD relates to the gap between what the child can learn without help and what he or she can learn with adult guidance or in collaboration with more capable peers.

Dynamic assessment embeds intervention within the assessment procedure. Typically there is a pre-test (to diagnose the characteristics of the student), an intervention (to provide appropriate feedback based on the
student’s characteristics and to help him or her learn), and then a post-test (to examine whether the student has learned the new concept after the intervention) (Haywood & Lidz, 2007). Despite the potential benefits of dynamic assessments, the need for experienced administrators and the amount of time involved have limited their use in applied educational settings (Frisby & Braden, 1992). Some researchers have reported the effectiveness of web-based dynamic assessments in enhancing students’ academic performance (Chen et al., 2010; Wang, 2010; Wang, 2011). Computerized dynamic assessments can be applied to the majority of students and would greatly improve the feasibility of using dynamic assessments in the classroom.

Based on the application of Vygotsky’s theory, Campione and Brown (1987) have developed the graduated prompt procedure to implement dynamic assessments. In the graduated prompt procedure, predetermined prompts that range from general to specific are given to provide students with gradual assistance until they solve the problem. In the beginning, students are provided general prompts. If students still cannot solve the problem after the first prompt, clearer and more specific prompts will be presented. The number of prompts required to solve the problem is taken as an indication of the students’ ZPD.

In the graduated prompt procedure, a student is given the same prompt (a general prompt) when he or she makes a mistake for the first time, regardless of which answer he or she chose. When the student makes a mistake for the second time, the student receive the same second level prompt (a specific prompt), regardless of which answer he or she chose. When the student makes a mistake for the third time, the teacher steps in and provides direct guidance (Campione & Brown, 1987). The graduated prompt procedure could promote precision in the estimation of individual learning ability and help teachers understand individual learning flexibility and modifiability (Hsu, 2008; Jitendra & Kameenui, 1993; Wang, 2011); however, little research has been done on the efficacy of the prompts in relation to students’ personal needs.

Computerized dynamic assessment can provide immediate feedback and is sensitive to progress. Providing feedback related to the process of problem solving could benefit students’ learning (Adesina, Stone, Batmaz, & Jones, 2014; Wang, 2010; 2011). A computerized adaptive test can offer customized items and reduce the length of the test and the time needed to take it. Adaptive dynamic assessments can benefit students’ learning in calculus (Ting & Kuo, 2015). Some preliminary studies on computerized dynamic adaptive test system had been introduced in the conference by Cheng, Wu, Kuo and Li (2013). The comprehensive results on evaluating the effectiveness of computerized dynamic adaptive test system in primary school Mathematics learning is presented in this study. The advantages of computerized dynamic assessment and adaptive tests are applied into a computerized dynamic adaptive assessment. The graduated prompt procedure that meets the students’ personal needs is integrated into the computerized dynamic adaptive assessment and will be referred to as “individual instruction prompts” in this paper.

**Aim of the study**

This study aimed to explore the effect of using the knowledge structure-based dynamic adaptive assessment as a remedial instruction program in comparison with a didactic teaching approach remedial instruction program in primary school mathematics. The research questions to be addressed in this study are as follows:

- How effective are the three remedial instruction programs at improving primary school mathematics?
- How effective is the computerized dynamic adaptive assessment compared with a didactic teaching approach?
- How efficient is the developed computerized dynamic assessment system for primary school mathematics?

**Domain experts’ knowledge structure**

In a task analysis, the cognitive skills required for successful learning and performance in a specific domain are usually defined by experts (Kalyuga, Rikers, & Paas, 2012). In this study, the “addition and subtraction of fractions with different denominators” unit of mathematics utilized in Taiwanese elementary schools is adopted by four practicing teachers and two domain experts (hereinafter referred to as “the experts”) to construct a knowledge structure. By analyzing teaching materials and objectives, the important cognitive skills of this unit were defined by the experts. After many discussions, the experts portrayed the sequence of the skills’ development and the relationships between these cognitive skills in a tree diagram. There were a total of eighteen cognitive skills for the “addition and subtraction of fractions with different denominators” unit. Figure 1 is an example of part of the experts’ knowledge structure, and it demonstrates that the upper-level cognitive skills, such as “apply addition and subtraction of fractions with different denominators,” are advanced cognitive skills,
while low-level cognitive skills, such as “calculate addition of improper fractions with different denominators,” are basic-level cognitive skills. “S” refers to the symbol of the cognitive skill. In this study, tests were developed based on the experts’ knowledge structure. Generally, an item was developed to assess knowledge on a single cognitive skill.

An adaptive test algorithm based on the knowledge structures suggested in the research of Wu et al. (2012) was used in this study. For example, S10 is forwardly linked to S11, which means that S10 must be mastered before S11 can be attempted; that is, S10 is a prerequisite to S11. If the student gets S11 correct, then it is inferred that he or she also knows its prerequisite (S10). This algorithm in computerized adaptive diagnostic tests can predict students’ learning profiles by utilizing fewer items than those used in original paper-based tests (Wu et al., 2012).

**Computerized dynamic adaptive test**

A computerized dynamic adaptive test system has been developed with PHP and MySQL on APACHE web servers. It consists of five modules: the Item Bank Management Module, Account Management Module, Test Management Module, Test Result Searching Module, and Dynamic Assessment Module. The Item Bank Management Module includes items and the knowledge structure of specific unit updates, modification, and management. The Account Management Module provides creation, modification, and management of user accounts. The Test Management Module sets the approach of test administration. Figure 2 represents the interface for an item’s administration. The Test Result Searching Module provides the group results (shown in Figure 3) and the individual results (shown in Figure 4) to the instructor and the student, respectively.

**Figure 1.** Part of the domain experts’ knowledge structure

**Figure 2.** The test administration interface

The Dynamic Assessment Module sets the algorithm of dynamic adaptive assessment based on the knowledge structure, as shown in Figure 5. First, the upper-level item with the most links is administered to a student. The status of the response is checked by the system. If the student answers the item correctly, then it is inferred that
the student also knows its prerequisites concepts. The item’s related prerequisite concepts will not be administered to the student and will be assumed to be answered correctly. The system checks whether or not there are any non-response items and, based on the results, either goes back to the first step or concludes the test. If the student answers an item incorrectly, based on the distracter option the student chose, the system will present the individualized instruction prompts with different levels to the student, as shown in Figure 6. A detailed description of the instruction prompts is given in the next section. If there are no more prompts, the process for solving this question will be provided to the student. Finally, the system checks whether or not there are any other non-response items and, based on the results, either goes back to the first step or concludes the test.

<table>
<thead>
<tr>
<th>Skill List</th>
<th>Passing Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test1 (%)</td>
<td>Test2 (%)</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Find equivalent fractions using expansion</td>
<td>97.6</td>
</tr>
<tr>
<td>Find equivalent fractions using reduction</td>
<td>88.1</td>
</tr>
<tr>
<td>Solve problems involving integers and fractions</td>
<td>59.5</td>
</tr>
<tr>
<td>Solve problems involving improper fractions and mixed fractions</td>
<td>81.0</td>
</tr>
<tr>
<td>Understand why fractions with different denominators are rewritten using common denominators the</td>
<td>97.6</td>
</tr>
<tr>
<td>Compare two fractions with different denominators by finding common</td>
<td>90.5</td>
</tr>
<tr>
<td>Solve problems involving adding proper fractions with different single-digit denominators</td>
<td>92.9</td>
</tr>
<tr>
<td>Solve problems involving adding proper fractions with different denominators by multiplying denominators with a single-digit number</td>
<td>59.5</td>
</tr>
<tr>
<td>Solve problems involving adding fractions with different denominators that require carrying</td>
<td>85.7</td>
</tr>
</tbody>
</table>

Figure 3. The profile for groups interface

**Individual Profile**

Dear Todd, it seems like you are not doing so well on the "addition and subtraction of fractions with different denominators" and there’s room for improvement which we’d love to see. So you need to spend more time working on this unit to make that happen. Please click the link below to start now.

| Basic Information |
|-------------------|----------------|
| Class             | 101            |
| Name              | Todd           |
| Location          | Taichung       |
| Test ID           | s10103         |
| Gender            | male           |

**Learning map**

1. O: master the skill ; X: not master the skill
2. The loading process might take you a few minutes. Please wait.

<table>
<thead>
<tr>
<th>Skill list</th>
<th>Diagnostic result</th>
<th>Tutorial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find equivalent fractions using expansion</td>
<td>O</td>
<td>tutorial 1</td>
</tr>
<tr>
<td>Find equivalent fractions using reduction</td>
<td>O</td>
<td>tutorial 2</td>
</tr>
<tr>
<td>Solve problems involving adding fractions with different denominators that require carrying</td>
<td>X</td>
<td>tutorial 9</td>
</tr>
<tr>
<td>Solve problems involving adding improper fractions with different denominators</td>
<td>X</td>
<td>tutorial 10</td>
</tr>
</tbody>
</table>
The experts designed the dynamic assessment test which consisted of 20 items, based on the experts’ knowledge structure shown in Figure 1 and the distracter options, based on the misconceptions, for the purpose of providing different instruction prompts for students who have different misconceptions related to the “addition and subtraction of fractions with different denominators” unit. When the student responded incorrectly to an item, he or she would be able to receive instruction prompts with different levels based on the number of times the student responded incorrectly for that item, thereby learning based on the nature of his or her misconceptions. In other words, each individual student would receive a different set of instruction prompts according to his or her needs.

**Instruction prompts**

The experts designed the dynamic assessment test which consisted of 20 items, based on the experts’ knowledge structure shown in Figure 1 and the distracter options, based on the misconceptions, for the purpose of providing different instruction prompts for students who have different misconceptions related to the “addition and subtraction of fractions with different denominators” unit. When the student responded incorrectly to an item, he or she would be able to receive instruction prompts with different levels based on the number of times the student responded incorrectly for that item, thereby learning based on the nature of his or her misconceptions. In other words, each individual student would receive a different set of instruction prompts according to his or her needs.
Taking item 1 in Table 1 as an example, the skill related to item 1 is “find equivalent fraction using expansion” and two misconceptions were embedded into the three distracter options. Based on the misconceptions that the student may have, individualized instruction prompts with different levels are provided by the system. Table 2 is an example of varied instruction prompts with different levels for item 1. When the student responds to item 1 incorrectly for the first time, a first level instruction prompt (general prompt) is given based on which distracter option was chosen. When the mistake is made for the second time, a second level instruction prompt (key word, specific prompt) is given based on the distracter option chosen. When the mistake is made for the third time, the third level instruction prompt, direct instruction is provided to the student.

**Table 1.** An example of developing an item according to skills and misconceptions

<table>
<thead>
<tr>
<th>Item 1</th>
<th>4/5 = (     )/40, what is the number in (     )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option</td>
<td>Related to skill or misconceptions</td>
</tr>
<tr>
<td>(1) 8</td>
<td>M1: when seeking an equivalent fraction, multiply the denominator and the numerator by different numbers</td>
</tr>
<tr>
<td>(2) 10</td>
<td>M1: when seeking an equivalent fraction, multiply the denominator and the numerator by different numbers</td>
</tr>
<tr>
<td>(3) 32</td>
<td>S1: find equivalent fraction using expansion</td>
</tr>
<tr>
<td>(4) 39</td>
<td>M3: falsely believe that adding a number to both the numerator and the denominator can result in an equivalent fraction</td>
</tr>
</tbody>
</table>

*Note. S1 and M1 are the symbols of cognitive skill and misconception, respectively.*

**Table 2.** Options-based instruction prompts of Item 1

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Option (1) 8</th>
<th>Option (2) 10</th>
<th>Option (3) 32</th>
<th>Option (4) 39</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without instruction prompt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First level prompt general prompt</td>
<td>Incorrectly! Your answer is 40 ÷ 5 = 8</td>
<td>Incorrectly! Your answer is 40 ÷ 4 = 10</td>
<td>Incorrectly! Your answer is 40 – 1 = 39</td>
<td>What is expansion?</td>
</tr>
<tr>
<td>Second level prompt keywords, concrete prompt</td>
<td>Incorrectly! The numerator is 5 × 8 = 40 How to expand the denominator?</td>
<td>Incorrectly! The numerator is 5 × 8 = 40 How to expand the denominator?</td>
<td>Incorrectly! Expansion of a fraction is not the numerator minus the denominator. How to expand the denominator?</td>
<td></td>
</tr>
<tr>
<td>Third level prompt direct instruction</td>
<td>The expansion refers to multiplying the numerator and denominator of a fraction by the same (bigger than 1) integer, which results in a fraction that is equivalent to the original fraction. Multiplying the numerator and denominator of $\frac{4}{5}$ by 8, that is, $\frac{4}{5} \times \frac{8}{8} = \frac{32}{40}$, the number in ( ) is 32.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Summative assessment**

Based on the experts’ knowledge structure, two summative assessments with alternate-form items were developed by the experts. The summative assessment, which consisted of 20 items, was used as the pre-test and post-test to evaluate the academic performance of students. The pre-test presents students’ understanding in the “addition and subtraction of fractions with different denominators” unit before the remedial instruction conducted in this study. The post-test demonstrates students’ understanding after receiving remedial instruction. Students were required to record the problem solving process in both their pre-test and post-test. The average difficulty index and the average discriminate index were 0.78 and 0.51, respectively. The Cronbach’s Alpha of the summative assessment was 0.93, indicating a high degree of internal reliability. The content validity for the summative assessment was reviewed by four domain experts (two professors and two elementary school mathematics teachers) with significant experiences in mathematics education. Items in the summative assessment did not show up in the dynamic assessment adaptive system.
Participants

One hundred and eighteen students (59 males and 59 females) from the fifth grade of a Taiwanese elementary school were selected to participate in this study. Using each class as a unit, students from the six classes were randomly assigned to three groups: a dynamic individualized assessment and individualized instruction group (DIA_II), an individualized instruction group (nDIA_II), and an instruction group (nDIA_nII). There were 42 students (18 males and 24 females) in the DIA_II group, 37 students (21 males and 16 females) in the nDIA_II group, and 39 students (20 males and 19 females) in the nDIA_nII group.

Experiment design

The pre-test/post-test nonequivalent group design was adopted in this study. Three remedial instruction programs were designed: the dynamic individualized assessment and individualized instruction group (DIA_II), the individualized instruction group (nDIA_II), and the instruction group (nDIA_nII). In the DIA_II group, the expert knowledge structure-based dynamic adaptive assessment was administered and students received the individualized instruction prompts based on the items and distracter options from the system. In the nDIA_II group, according to the pre-test results, the individualized remedial instruction based on the expert knowledge structure was provided by the system. In the nDIA_nII group, the group remedial instruction based on the classroom report of the pre-test was performed by the teachers.

The difference in experimental treatment between the three groups is shown in Table 3. Both the DIA_II and nDIA_II groups received computerized individualized instruction based on the expert knowledge structure. The students in the DIA_II group received prompts according to the options they chose and the students in the nDIA_II group received direct instructions. In the nDIA_nII group, cognitive skills were taught by the teachers in sequence, based on the group report of the pre-test. Since cognitive skills with a lower correct response rate indicated relative weaknesses for a specific group, these cognitive skills were targeted as a priority for remedial instruction. The pedagogy for the nDIA_nII group was more like a traditional remedial teaching program (i.e., didactic teaching) commonly used in Taiwan.

<table>
<thead>
<tr>
<th>Table 3. Research design of the three groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIA_II</td>
</tr>
<tr>
<td>Individualized prompt</td>
</tr>
<tr>
<td>Individualized instruction</td>
</tr>
</tbody>
</table>

Experiment procedure

The experiment procedure is shown in Figure 7.
First, all students received five periods of mathematics instruction for an “addition and subtraction of fractions with different denominators” unit. One period is 40 minutes. The curriculum used for the three groups was the same. After mathematics instruction, all students took the pre-test of the summative assessment. The function of this pre-test was to understand the learning conditions of the students prior to the remedial instruction. The six classes were randomly assigned to three groups and students took part in one of the three remedial instruction programs. Finally, to understand the effectiveness of the remedial instruction programs, all students took the post-test of the summative assessment. The control factors were grade five students, instruction time, pre-test time, experiment time, post-test time, and instruction material.

Data collection and analysis

The data collected in this research included the pre-test scores of the summative assessment and the post-test scores of the summative assessment. The utilization of test items and accuracy rate were computed as criteria to evaluate the efficiency of the knowledge structure-based dynamic adaptive assessment system. While the students completed the adaptive portion of the test, the system continued to administer the rest of the 20 items in order to compute the accuracy rate. The utilization of test items is determined by the average number of items administered to the examinees in the knowledge structure-based dynamic adaptive assessment system. The mathematics formula of accuracy rate is (the total number of items—the items which inferred answered correctly but failed) / the total number of items.

The study used PASW statistics 18 (SPSS, Inc.) to conduct data analysis. To explore the performance of the three remedial instruction programs (DIA_II, nDIA_II, and nDIA_nII) in improving learning, a paired-samples t-test was conducted to compare the scores in pre-test and post-test conditions. In addition, analysis of covariance (ANCOVA) was used to detect significant differences between the three remedial instruction programs. In ANCOVA, the pre-test scores were taken as the covariate, the post-test scores were taken as the dependent variable, and the three remedial instruction programs were taken as the fixed factor. The Least Significant Difference (LSD) method was used to perform post hoc analysis. The criterion for the level of statistical significance was set at 0.05.

Effectiveness of the three remedial instruction programs

In order to investigate whether students showed an improvement after participating in the remedial instruction programs, a paired sample t-test was carried out, as shown in Table 4. The difference in mean scores between pre-test and post-test was 12.14 for the dynamic individualized assessment and individualized instruction group (DIA_II). The difference in mean scores between pre-test and post-test was 8.51 for the individualized instruction group (nDIA_II). The difference in mean scores between pre-test and post-test was 7.95 for the instruction group (nDIA_nII). As Table 4 shows, the average score of the post-test was significantly higher than the average pre-test score (\( p = .000 < .05 \)) for all three groups, which demonstrates that the students showed an improvement after the remedial instruction programs.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-test score</th>
<th>Post-test score</th>
<th>( t )-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>DIA_II (n = 42)</td>
<td>80.12</td>
<td>15.87</td>
<td>92.26</td>
</tr>
<tr>
<td>nDIA_II (n = 37)</td>
<td>81.22</td>
<td>13.61</td>
<td>89.73</td>
</tr>
<tr>
<td>nDIA_nII (n = 39)</td>
<td>81.67</td>
<td>13.39</td>
<td>89.62</td>
</tr>
</tbody>
</table>

Table 4. Results of difference in scores between pre-test and post-test for three groups

Effectiveness of the computerized dynamic adaptive assessment

ANCOVA was conducted to investigate the differences in learning effectiveness of the different remedial instruction programs. Homogeneity of variance was conducted first; the result of Levene’s test showed the requirement of homogeneity of variance was not violated \( (F(2,115) = 1.11; p = .332) \). In addition, the assumption for homogeneity of regression coefficients was also conducted \( (F(2,115) = 0.058; p = .944) \). That is, the homogeneity for regression coefficients within the three groups is confirmed.
The results of ANCOVA are shown in Table 5, which reveals the effect of different remedial instruction programs on post-test scores after adjusting for the effect of the pre-test scores. According to Table 5, there were significant differences in post-test scores between the three types of remedial instruction programs. The results of the LSD test, which showed that the difference between the dynamic individualized assessment and individualized instruction group (DIA_II) and the individualized instruction group (nDIA_II), was 3.23 (p < .05), which is statistically significant. The difference between the dynamic individualized assessment and individualized instruction group (DIA_II) and the instruction group (nDIA_nII) was 3.64 (p < .05), which is also statistically significant. However, the difference (0.40) between the nDIA_II group and the nDIA_nII group was not significant (0.40) (p > .05).

The results indicated that the DIA_II group outperformed the nDIA_II group and nDIA_nII group. The knowledge structure-based dynamic adaptive assessment which provides dynamic individualized assessment and individualized instruction could effectively benefit mathematics students. Wang (2011) pointed out that graded prompts in the assessment in comparison to direct assessment (web-based assessment and paper-pencil assessment) can help junior high school students learn mathematics. In this study, compared to direct instruction (nDIA_II and nDIA_nII), the graduated prompts based on options could not only provide individualized instruction to meet the students’ needs but could also assist elementary school students in learning mathematics.

The instruction based on the expert knowledge structure was provided to both groups (nDIA_II and nDIA_nII). The students in the nDIA_II group were taught by computers, based on an individualized profile, and the students in the nDIA_nII group were taught by teachers using a group profile. There were no significant differences between the nDIA_II group and the nDIA_nII group, which demonstrates that the results from technological assistance were as good as those from teachers delivering remedial instruction. The primary advantage of technological assistance was that it reduced human resources costs.

### Table 5. Results of ANCOVA on the learning effectiveness of the three remedial instructions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level</th>
<th>Mean± (SE)</th>
<th>F values</th>
<th>Post hoc&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td>250.137&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>DTRIP</td>
<td>DIA_II</td>
<td>92.81</td>
<td>4.154&lt;sup&gt;a&lt;/sup&gt;</td>
<td>DIA_II &gt; nDIA_II&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>nDIA_II</td>
<td>89.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nDIA_nII</td>
<td>89.17</td>
<td></td>
<td>DIA_II &gt; nDIA_nII&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Note. <sup>a</sup>p < .05. DTRIP: Different Types of Remedial Instruction Programs. <sup>b</sup>Covariates appearing in the model are evaluated at the following value: 80.97. <sup>a</sup>Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

### The pre-test and post-test passing rate on cognitive skills

A deeper discussion on the dynamic individualized assessment and individualized instruction group via passing rates on cognitive skills is provided in this section. The pre-test and post-test were developed based on the expert knowledge structure consisting of eighteen cognitive skills. The passing rate on cognitive skills is the rate of questions answered correctly. For example, if question 1 measures skill 1, the passing rate on cognitive skill 1 is 97.62% (the percentage of students who answered question 1 correctly). In this study, skills were measured by one question, except cognitive skill 11 and cognitive skill 18, which were measured by two questions each. The passing rate on cognitive skills is shown in Table 6. Since the pre-test was used to diagnose whether a student is equipped with the cognitive skills for the “addition and subtraction of fractions with different denominators” unit and provided feedback to the remedial instruction programs, the easier questions were developed by the teachers. Therefore, the passing rate is rather high for most cognitive skills except for cognitive skill 18. As a result, the improvement that can be demonstrated is relatively low for some cognitive skills such as skill 2, skill 11, and skill 14.

Table 6 demonstrates that student performance for cognitive skills 5 and 17 showed a decline after the dynamic individualized assessment and individualized instruction. The reason for this may be that cognitive skill 5 is measured by a concept problem where no calculation step is required and the choices were changed for the post-test, so the students may have been too careless. Cognitive skill 17 is measured by an item without a wording context, 3 -2(7/15) = ( ). A few students mistook the problem on subtraction as a problem on addition, which resulted in an error. More notably is the highest improvement rate on cognitive skill 18 (32.14%), which is the most upper-level cognitive skill in the expert knowledge structure. In other words, cognitive skill 18 is the hardest cognitive skill in the expert knowledge structure. This means that the dynamic individualized assessment
and individualized instruction program can help students to improve their academic performance, especially when learning harder cognitive skills.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Improvement</th>
<th>Skill</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>97.62</td>
<td>100</td>
<td>2.38</td>
<td>S10</td>
<td>83.33</td>
<td>97.62</td>
<td>14.29</td>
</tr>
<tr>
<td>S2</td>
<td>88.10</td>
<td>97.62</td>
<td>9.52</td>
<td>S11</td>
<td>92.86</td>
<td>97.62</td>
<td>4.76</td>
</tr>
<tr>
<td>S3</td>
<td>59.92</td>
<td>83.33</td>
<td>23.81</td>
<td>S12</td>
<td>78.57</td>
<td>92.86</td>
<td>14.29</td>
</tr>
<tr>
<td>S4</td>
<td>80.95</td>
<td>88.10</td>
<td>7.15</td>
<td>S13</td>
<td>80.95</td>
<td>83.33</td>
<td>2.38</td>
</tr>
<tr>
<td>S5</td>
<td>97.62</td>
<td>92.86</td>
<td>-4.76</td>
<td>S14</td>
<td>80.95</td>
<td>100</td>
<td>19.05</td>
</tr>
<tr>
<td>S6</td>
<td>90.48</td>
<td>100</td>
<td>9.52</td>
<td>S15</td>
<td>85.71</td>
<td>97.62</td>
<td>11.91</td>
</tr>
<tr>
<td>S7</td>
<td>92.86</td>
<td>100</td>
<td>7.14</td>
<td>S16</td>
<td>64.29</td>
<td>83.33</td>
<td>19.04</td>
</tr>
<tr>
<td>S8</td>
<td>59.52</td>
<td>88.10</td>
<td>28.58</td>
<td>S17</td>
<td>97.62</td>
<td>95.21</td>
<td>2.38</td>
</tr>
<tr>
<td>S9</td>
<td>85.71</td>
<td>92.86</td>
<td>7.15</td>
<td>S18</td>
<td>46.43</td>
<td>78.57</td>
<td>32.14</td>
</tr>
</tbody>
</table>

The efficiency of the computerized dynamic adaptive assessment

Two criteria, the utilization of test items and accuracy rate, were used to evaluate the efficiency of the knowledge structure-based dynamic assessment system. This test consisted of 20 items. The average utilization of test items was 3.92. In other words, the system could reduce approximately 16 test items on average. The number of inferred answered correctly but failed items average was 2.17; therefore, the accuracy rate was 89.2 % ((20-2.17)/20).

The experts designed the dynamic assessment test which consisted of 20 items, based on the experts’ knowledge structure (as described in Figure1) and the distracter options, based on the misconceptions, for the purpose of providing different instruction prompts for students who have different misconceptions related to the “addition and subtraction of fractions with different denominators” unit. When the student responded incorrectly to an item, he or she would be able to receive instruction prompts with different levels based on the number of times the student responded incorrectly for that item, thereby learning based on the nature of his or her misconceptions. In other words, each individual student would receive a different set of instruction prompts according to his or her needs.

This result was acceptable according to the research results of Wu et al. (2012). The knowledge structure-based dynamic adaptive assessment provided individualized prompts, which implied instruction, so the learning of a student was based on the nature of his or her misconceptions. Based on the knowledge structures adaptive test algorithms, if the student responds to an item correctly, it is inferred that he or she also knows its prerequisites; therefore, the average utilization of test items was low in this research. The knowledge structure was defined by the experts in this research, and a differently constructed knowledge structure approach could be attempted to increase the accuracy rate (Wu et al., 2012).

Conclusion

Dynamic assessment is sensitive enough to provide information about an individual’s ability, change processes, and the mediation strategies responsible for cognitive modifiability. However, dynamic assessment takes more time and requires more skill than static testing. A computerized adaptive test can increase the efficiency of the testing process by offering customized items. Nevertheless, a computerized adaptive test is not sensitive to progress within a test. The computerized dynamic adaptive assessment system which combines the advantages of both dynamic assessment and computerized adaptive tests was developed in this study. The selling point of the computerized dynamic adaptive assessment system is that it provides both tailored assessment and individualized prompting. Three remedial instruction programs – the dynamic individualized assessment and individualized instruction, individualized instruction, and didactic teaching – can improve elementary school students’ performance in mathematics.

The students with tailored assessment and individualized prompting score significantly higher in primary school Mathematics than those with individualized instruction and with didactic teaching, respectively. The computerized dynamic adaptive assessment system can help students learning effectively. Although this study has offered valuable insights into dynamic assessment and pedagogical implications, its design is not without flaws. The first limitation concerns the knowledge structure defined by the experts. The risk of constructing
incorrect links between cognitive skills will lead to a decrease in the accuracy rate of this study. The second limitation is rooted in guessing correctly. Further research could incorporate a constructed response item into the knowledge structure-based dynamic adaptive assessment system. Finally, the generalization of the results to other content with different populations may be limited.

Acknowledgements

The authors would like to thank the Ministry of Science and Technology in Taiwan (Grant No. MOST 103-2511-S-656-003-MY2 and 102-2511-S-142 -008 -MY3) for partly supporting this study.

References


Comment Data Mining to Estimate Student Performance Considering Consecutive Lessons

Shaymaa E. Sorour1,2*, Kazumasa Goda3 and Tsunenori Mine4

1Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan // 2Faculty of Specific Education, Kaft Elsheikh University, Egypt // 3Kyushu Institute of Information Science, Japan // 4Faculty of Information Science and Electrical Engineering, Kyushu University, Japan // shaymaa@ma.ait.kyushuu.ac.jp // gouda@kiis.ac.jp // mine@ait.kyushu-u.ac.jp

*Corresponding author

(Submitted October 22, 2015; Revised March 1, 2016; Accepted May 17, 2016)

ABSTRACT

The purpose of this study is to examine different formats of comment data to predict student performance. Having students write comment data after every lesson can reflect students’ learning attitudes, tendencies and learning activities involved with the lesson. In this research, Latent Dirichlet Allocation (LDA) and Probabilistic Latent Semantic Analysis (pLSA) are employed to predict student grades in each lesson. In order to obtain further improvement of prediction results, a majority vote method is applied to the predicted results obtained in consecutive lessons. The research findings show that our proposed method continuously tracked student learning situations and improved prediction performance of final student grades.

Keywords

Learning analytics, Free-style comments, Topic models, Majority vote

Introduction

Nowadays, higher education institutions play a crucial role in improving the quality and the efficacy of education. Thus, it is important to determine methods that can provide quality education to students. One way to enhance the quality of educational processes is to focus on learners’ activities and attitudes, which are currently unknown to the decision makers and very valuable in analyzing students’ behavior, predicting performance and assisting courseware authors in detecting shortcomings and providing feedback to improve student performance. However, it is a difficult task to predict students’ performance by analyzing their behavior because each student has his/her own characteristics that influence their performance, such as their demographic, cultural, social, or psychological profile, previous schooling, prior academic performance, interactions between students and faculty, etc. (Yadav & Pal, 2012; Chalaris, Gritzalis, Maragoudakis, Sgouropoulou, & Tsolakidis, 2014; Natek & Zwilling, 2014).

Comment data can potentially eliminate barriers between students and their teacher, and provide opportunities so that students can think back on their learning behaviors in connection with the subject. Thus, it becomes an essential accessory to support their learning. Further, comment data enables students to express themselves, their attitudes, and interactions, and to reflect their learning activities and difficulties for each lesson, especially for those having introverted characters and who are not comfortable expressing their views or asking questions. In addition, it allows teachers at the same time to improve their way of giving lessons, and of contacting students.

In this paper, we synthesize the learning analytics (LA) and data mining (DM) approaches to explore the development of more usable prediction models of student performance using free-style comment data written by students after every lesson. The ability to predict the final performance of a student has been growing steadily in education. There are specific student characteristics which can be associated with the student success rate. One of the practical applications of student performance prediction is to find relationships between the work done by each student and the precise mark obtained. If we could predict student performance in advance, a feedback process could help to improve the learning process of the students during the course (Natek & Zwilling, 2014; Zafr & Ventura, 2009; Xing, Guo, Petakovic, & Goggins, 2015).

To further contribute to the understanding of individual students in the class, this paper presents a study that examines different types of comment data (i) to find students’ situations, tendencies and attitudes by predicting final student grades, (ii) to achieve further improvement in predicting final student grades by considering the student grades predicted in consecutive lessons and to keep track of the student learning situation, and (iii) to examine the characteristics of comment data collected from three different classes affecting the accuracy of final student grade prediction.
The objective of our study is to find a reply to the following research questions:

**Question 1.** Which topic model has the best results in predicting student performance, pLSA or LDA?
**Question 2.** What comment format is the best predictor?
**Question 3.** Is it possible to make an accurate prediction in early lessons, such as by lesson five?
**Question 4.** Is there any relation between the difficulty of a subject and accuracy in predicting final student grades?

The rest of the paper is organized as follows. Related work gives an overview of some related literature. Preliminary introduces an overview of our research and the procedures of the proposed methods. Methodology describes our proposed methods. Results discusses some of the highlighted experimental results. Finally, conclusions and future work concludes the paper and describes our future work.

**Related work**

Predicting student performance is one of the most useful applications of Educational Data Mining (EDM) and LA; its goal is to estimate student performance, knowledge and score from other information and aspects or behavior of the students (Romero & Ventura, 2013; Papamitsiou & Economides, 2014). Within the intelligent tutoring systems research community, there has been substantial work on constructing student models that can accurately predict student performance (e.g., Romero, López, Luna, & Ventura, 2013; Zafra & Ventura, 2009; He, 2013; Khajah, Wing, Lindsey, & Mozer, 2014; Pavlik, Cen, & Koedinger, 2009). An emerging trend in EDM is the use of text mining. Text mining focuses on finding and extracting useful or interesting patterns and models from unstructured text documents. As an automated technique, text mining can efficiently be used to identify, extract, integrate, and exploit knowledge for research and education (He, 2013; Ananiadou, 2007).

Currently, there are only several studies using text mining techniques to analyze learning-related data. (Abdous, He, & Yen, 2012; He, 2013) tried to correlate interactions within the online questions and chat messages automatically recorded by a live video streaming environment to predict final student grades and understand the ways students are engaged in online activities. In addition, (Minami & Ohura, 2013) analyzed student attitudes towards learning, and investigated how they had affected their final evaluation; they pursued a case study of lecture data analysis in which there are correlations between student attitudes to learning, such as attendance and homework, as effort, and the student examination scores, as achievement. (Şen, Uçar, & Delen, 2012) identified the factors that lead students to success or failure in placement tests. They developed models to predict secondary education placement test results and used sensitivity analysis on those prediction models to identify the most important predictors. Further, (Minami & Ohura, 2015) investigated correlations of students’ achievement and their learning characteristics by analyzing the texts of their answers to a questionnaire. The results showed different words usage distribution between low, middle and high achievement student. (Goda & Mine, 2011) proposed the PCN method to estimate learning situations from comments freely written by students after every lesson. The PCN method asks students to describe their learning attitudes or behaviors in a format consisting of three viewpoints: P (Previous activity), C (Current activity), and N (Next activity); students’ comments can be analyzed from the point of view of their time-oriented learning situations. For the purpose of predicting final student grades from comment data, (Sorour, Mine, Goda, & Hirokawa, 2015) proposed the LSA and k-means clustering method, and the overlap and the similarity measuring methods. (Sorour, Goda, & Mine, 2015) also applied two types of machine learning techniques: ANN and SVM, and made them learn the relationships between comment data analyzed by LSA and final student grades. The average prediction F-measure results of final student grades were 78.8% and 84.2% for ANN and SVM, respectively.

In this study, new methods are proposed for predicting student performance with high accuracy using different formats of comment data with the same PCN method. The majority vote method is employed to keep track of each student’s learning situation based on consecutive multiple lessons.

**Preliminary**

**Comment data and subject of the study**

In our research, we used two types of comment data collected from Japanese students who attended Goda’s and Mine’s courses.
**Goda’s course**

The second author collected comments from students who attended the Information Processing course. There were two classes: (Class A = 60 students) and (Class B = 63 students). Students took the same subject. It consisted of 15 lessons (one a week). The main subject from lessons 1 to 6 was Computer Literacy, giving information on how to use some IT tools. Computer Literacy education is compulsory throughout senior high schools in Japan, with only a few differences in the details of course contents. From lessons 7 to 15, students learned the Introduction of C Programming course. It is a new subject and not required until they enter the university. Class A students from School of Pharmacy and Class B students from School of Agriculture.

Each student wrote his/her comments concerning each lesson from the three viewpoints: P, C, and N. The teacher asked students to fill in three simple items about their learning status. As a result, we collected about 2600 short comments for each class. We call these comments “Comments from classes A and B.” If there is no ambiguity, we just call them Class A and Class B. Table 1 summarizes examples of the three viewpoints for classifying comments.

**Mine’s course**

The third author collected comment data from one class. It consisted of 89 students from School of Engineering. The main course was an Introduction of C Programming. Most of the students were novices at programming. The aim was to collect comment data with high quality and to increase students’ motivation to express their learning activities and attitudes more deeply after each lecture. Comment data was collected using the same PCN method, but we employed a format consisting of the five questions shown in Table 1. We call these comments Comments from Class C or just Class C if there is no ambiguity.

<table>
<thead>
<tr>
<th>Table 1. Examples of comment data</th>
<th>Goda’s course</th>
<th>Mine’s course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>Goda’s course</td>
<td>Mine’s course</td>
</tr>
<tr>
<td>P Learning activity before the class time.</td>
<td>I have logged into WebCT at home and prepared lessons according to the materials before class.</td>
<td>Q1: What did you do for this lecture? How long did you study for this lecture?</td>
</tr>
<tr>
<td>C Understanding and achievements during the class time.</td>
<td>I have tried to learn and practice a programming language.</td>
<td>Q2: Do you have anything you did not understand? Any questions?</td>
</tr>
<tr>
<td>C Q3: What did you solely find in this lecture?</td>
<td>I found the way for rounding of the decimal point.</td>
<td>Q4: Did you teach something to your friends to, get answers or discuss with them?</td>
</tr>
<tr>
<td>N Learning activity plan until the next class.</td>
<td>It will be even harder in the following lessons; I will keep doing a preview before classes.</td>
<td>Q5: What will you plan to do for the next lecture?</td>
</tr>
</tbody>
</table>

Compared to the PCN items, (Q1) corresponds to a P-item and (Q5) is an N-item. (Q2), (Q3), and (Q4) are related to C-items, although they were not distinguished in Class A and B. In (Q2) students wrote whether they understood the lesson, what parts they did not understand and why. In (Q3), students described what they found and realized in the lesson after making the efforts and to what extent they made efforts. In (Q4), students described if there is any collaboration with their friends such as discussion, teaching or solving a problem with each other. From the previous questions and teacher’s instructions, students can describe their comments with high quality compared to the normal PCN method (Goda & Mine, 2011). Further, all students wrote about their attitudes, understanding, willing to improve the level and overcome problems.
We consider predicting each student’s performance from his/her comments. We choose five grades instead of the mark itself as a student’s result. Table 2 shows the correspondence between the grades and the mark for the three class data sets.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Mark</th>
<th>Class A</th>
<th>Class B</th>
<th>Total</th>
<th>Class C</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>100-90</td>
<td>11</td>
<td>10</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>A</td>
<td>89-80</td>
<td>17</td>
<td>24</td>
<td>41</td>
<td>31</td>
</tr>
<tr>
<td>B</td>
<td>79-70</td>
<td>17</td>
<td>6</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>C</td>
<td>69-60</td>
<td>8</td>
<td>9</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>D</td>
<td>59-0</td>
<td>7</td>
<td>14</td>
<td>21</td>
<td>4</td>
</tr>
</tbody>
</table>

The assessment of each student was established by considering the average mark of his/her reports of three assignments about IT tools/programming during the period of the semester, and his/her attendance rate. In class C only four students achieved a D grade, and there were two students who did not submit comments for most of the lessons. So, we combined comment data from grade D into grade C.

Methodology

This work proposes methods for predicting final student performance based on different types of comment data. The overall procedure of our proposed methods shown in Figure 1 is as follows:

Comment data collection

Student comments are collected after each lesson. Teachers ask students to send their comments by e-mail, or to write and submit their comments on the back of an attendance card. In our study, two teachers collected comments from three classes.

Data preparation

Our method uses the Japanese morphological analyzer MeCab to analyze comments and to extract words and parts of speech (Kudo, 2013). In our study, we only used nouns, verbs, adjectives, and adverbs. The number of words appearing in comments in each of A and B classes was about 780 words (over 310 words without duplication). The number of words appearing in each comment averaged 14. In class C, the number of words appearing in each comment and that in all comments for each lesson averaged 12 and 1,115, respectively. The number of distinct words in all the comments for each lesson was over 520.

Next, we create a word-by-comment matrix A with extracted words. Matrix A is comprised of \(m\) words \(w_1, w_2, \ldots, w_m\) in \(n\) comments \(c_1, c_2, \ldots, c_n\), where the value of each cell \(a_{ij}\) indicates the total occurrence frequencies of word \(w_j\) in comment \(c_i\); we apply pLSA and LDA models to the matrix A to generate relationships between the extracted words and latent concepts contained in the unstructured collection of texts (students’ comments).

Topic models

The primary goal of this research is to find the best model to treat comment data while preserving the essential statistical relationships that are useful for basic tasks such as classification or prediction. pLSA and LDA have all been demonstrated to be reliable methods in Information Retrieval (IR) (Hofmann, 1999; Blei et al., 2003).

Training phase

This phase builds prediction models of student grades based on pLSA and LDA results as input data. The SVM model is employed to build strong relationships between the input data and a student grade, which is the output data.
Test phase

This phase evaluates the performance of prediction models by calculating accuracy and F-measure for each lesson and estimates each student’s performance in consecutive lessons using the majority vote method.

Topic models

pLSA provides a probabilistic formulation to model documents in a text collection. It assumes that the words are generated from a mixture of latent aspects (topics) which can be from a document (Hofmann, 1999). There is a hidden (latent) topic variable $z_k$ associated with the occurrence of a word $w_j$ in a comment $c_i$.

The joint probability $P(w_j, c_i, z_k)$ is assumed to have the form of the graphical model as shown in Figure 2(a). We can obtain the conditional probability $P(w_j, c_i)$ by marginalizing over the latent aspect $z_k$

$$P(w_j, c_i) = P(c_i) \sum_{k=1}^{K} P(z_k | c_i) P(w_j | z_k)$$  \hspace{1cm} (1)$$

where $P(z_k | c_i)$ is the probability of aspect $z_k$ occurring in a comment $c_i$ and $P(w_j | z_k)$ is the probability of word $w_j$ occurring in a particular aspect $z_k$. Fitting the model to the observed data can be completed using the well-known Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). The EM algorithm maximizes the likelihood of observed data by estimating the parameters $P(w_j | z_k)$ and $P(z_k | c_i)$.

$$L = \prod_{i=1}^{m} \prod_{k=1}^{K} P(c_i) \sum_{k=1}^{K} P(z_k | c_i) P(w_j | z_k)^{n(w_j, c_i)}$$  \hspace{1cm} (2)$$

Figure 1. A framework of the methodology

Figure 2. Graphical model of pLSA/LDA
where \( n(w_j, c_i) \) denotes the count of words \( w_j \) in comment \( c_i \). Maximizing the log of likelihood \( L \) in Equation (2) with respect to the probabilistic distribution, the iteration equations can be obtained using the E-Step and M-Step of the EM algorithm (Hofmann, 2001).

LDA treats the multinomial weights over topics as latent random variables (Blei, Ng, & Jordan, 2003). It allows the assignment of probabilities to data outside the training corpus and uses fewer parameters, thus reducing overfitting. The LDA model is shown in Figure 2(b). The goal is to maximize the following likelihood:

\[
P(w | \Phi, \alpha, \beta) = \prod_c \sum_z P(w | z, \Phi) P(z | \theta) P(\theta | \alpha) P(\Phi | \beta)\]

where \( \theta \) and \( \phi \) are multinomial parameters over the topics and words, and \( P(\theta | \alpha) \) and \( P(\phi | \beta) \) are dirichlet distributions parameterized by the hyper-parameters \( \alpha \) and \( \beta \). Since the integral is intractable to solve directly, Gibbs sampling is often used to solve the parameters (Griffiths & Steyvers, 2004).

In this paper, we compare the LDA model with the pLSA model by examining the optimal number of topics and words that scored high prediction results.

**How many topics?**

The main difficulty in our application of the pLSA/LDA models is to choose the optimal number of topics in the training phase so as to get the best prediction results. In our research, we checked the average F-measure \( F_1 \) of prediction results using the different numbers of topics with the SVM model. We found the highest F-measure results by varying the number of topics from 2 to 30 as reducing the size of topics.

**Topic interpretation**

In this research, we employed topic models to reveal and interpret the latent intellectual topics in students’ comments. Topics listed in Table 3 from the LDA model and C-comments are readily recognizable as they are related to students’ situations in the classroom. Students in Class A and B described their learning activities and topics related to the current lesson more than illustrating their attitudes towards the lesson subject. On the other hand, students in Class C clearly wrote about their attitudes using positive or negative words and described different situations (e.g., understanding, collaboration, attitudes), besides topics and activities related to each lesson. We manually sorted these topics into some categories, as shown in Table 4. Using these topics, we can
easily distinguish the upper and lower grades of Class C students. Upper grade students often wrote about their finding. They described the points that they found and realized with their efforts. Also, they had positive attitudes toward the lesson and collaborated with each other in most lessons. On the other hand, lower grade students did not understand the first and the last lessons. They tended to have negative attitudes toward the lessons. In a few lessons, they did not collaborate with their friends and did not find new points with their efforts.

Table 3. Top 10 high-frequency words for each topic in the 8-topic model

<table>
<thead>
<tr>
<th>Topic</th>
<th>High-frequent words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A and B</td>
<td></td>
</tr>
<tr>
<td>Topic 1</td>
<td>Program, email, information, RAM, Excel, PC, graph, power point, tool, website.</td>
</tr>
<tr>
<td>Topic 2</td>
<td>Review, remember, listen, study, understand, finish, fast, think, quickly, end.</td>
</tr>
<tr>
<td>Topic 3</td>
<td>Submit, early, level, work hard, browse, search, exit, edit, type, write.</td>
</tr>
<tr>
<td>Topic 4</td>
<td>Study, use, learn, calculate, practice, information, think, convert, type, design.</td>
</tr>
<tr>
<td>Topic 5</td>
<td>Arrow, type, book, cell, program, unit, complete, SMTP, time, IMAP.</td>
</tr>
<tr>
<td>Topic 6</td>
<td>Perform, create, exercise, sheet, function, number, case, size, design, calculate.</td>
</tr>
<tr>
<td>Topic 7</td>
<td>Positive, keep up, easy, like, capable of, interesting, fix, discover, solve, repair.</td>
</tr>
<tr>
<td>Topic 8</td>
<td>Strong, difficult, hard, worry, mistake, frustrate, complex, bad, lost, boring.</td>
</tr>
</tbody>
</table>

Class C

<table>
<thead>
<tr>
<th>Topic</th>
<th>High-frequent words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td>Understand, learn, know well, create, treat, easy, practice, complete, fix.</td>
</tr>
<tr>
<td>Topic 2</td>
<td>Found, function, realize, discover, conclude, examine, decimal, solve, fix, show.</td>
</tr>
<tr>
<td>Topic 3</td>
<td>Characters, function, array, method, string, assign, swap, program, calculate, knowledge.</td>
</tr>
<tr>
<td>Topic 4</td>
<td>Friend, tell, discuss, talk, ask, cooperate, help, taught, solve, consultation.</td>
</tr>
<tr>
<td>Topic 5</td>
<td>Worry, difficult, complex, error, create, treat, fix, practice, message, forgot.</td>
</tr>
<tr>
<td>Topic 6</td>
<td>Hard, unable to, not well, confuse, can’t be, fail, sad, wrong, anxiety, nervous.</td>
</tr>
<tr>
<td>Topic 7</td>
<td>Examine, finish, try, work hard, practice, create, concentrate, hope, handle well, cooperate.</td>
</tr>
<tr>
<td>Topic 8</td>
<td>Able to, avoid, do the best, satisfy, helpful, good, like, fun, interest, can be.</td>
</tr>
</tbody>
</table>

Table 4. Categorization of the 8 topics

<table>
<thead>
<tr>
<th>Category</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A and B</td>
<td></td>
</tr>
<tr>
<td>Topics in the current lesson</td>
<td>Topics 1, 5</td>
</tr>
<tr>
<td>Learning rate and student behavior</td>
<td>Topics 2, 3</td>
</tr>
<tr>
<td>Activities</td>
<td>Topics 4, 6</td>
</tr>
<tr>
<td>Student learning attitudes</td>
<td>Topics 7, 8</td>
</tr>
<tr>
<td>Class C</td>
<td></td>
</tr>
<tr>
<td>Understanding the lesson subject</td>
<td>Topic 1</td>
</tr>
<tr>
<td>Finding in the current lesson</td>
<td>Topic 2</td>
</tr>
<tr>
<td>Collaboration with friends</td>
<td>Topic 4</td>
</tr>
<tr>
<td>Student learning attitudes</td>
<td>Topics 5, 6, 8</td>
</tr>
<tr>
<td>Learning rate and student behavior</td>
<td>Topics 3, 7</td>
</tr>
</tbody>
</table>

Model estimation by SVM

SVM is a powerful solution to the classification problems. In our research, we have some pretreatments to prepare training data for SVM. From the previous phase, we have got a number of topics and their corresponding words and probabilities. Then, we find out all the words which one student used in his/her comment and create a topic vector according to the distinct words collected in the student’s comment. After obtaining a list of topic vectors of all students, we used the MATLAB LibSVM tool (Chang & Lin, 2011) to predict final student grades. We employed the SVM model with a radial basis function (RBF) kernel to generate models from lessons 1 to 15 and predict a student grade as one of five grades: S, A, B, C and D, based on the analyzed comment data.

Measure of prediction performance

Evaluation criteria

In our experiments, the prediction performance of the proposed models was evaluated by 10-fold cross validation using the comments from the three classes.
Effectiveness for text classification was measured by two different means: F-measure (F1) and accuracy (Acc). F-measure is the most important metric, which is equal to the harmonic mean of Recall (R) and Precision (P) (Yang & Liu, 1999).

Let \( l_i \) be the \( i^{th} \) lesson, where \( 1 \leq i \leq 15 \). Let \( G = \{ S, A, B, C, D \} \) be the set of grades. The overall \( F_1 \) score of the entire classification problem can be computed by macro-average (\( F_{1}^{\text{ME}} \)).

In macro-averaging (\( F_{1}^{\text{ME}} \)), F-measure for each lesson \( l_i \) is computed locally over each grade \( g \) first and then the average over all grades \( G \) is taken. \( R \) and \( P \) are computed for each grade as in Equation 4. Then \( F_1 \) for each grade is computed and \( F_{1}^{\text{ME}} \) is obtained by taking the average of \( F_1 \) values for each grade in Equation 5:

\[
P_g = \frac{TP_g}{TP_g + FP_g}, \quad R_g = \frac{TP_g}{TP_g + FN_g}
\]

\[
F_{1g} = 2 \cdot \frac{P_g \cdot R_g}{R_g + P_g}, \quad F_{1}^{\text{ME}} = \frac{\sum_{g \in G} F_{1g}}{|G|}
\]

The accuracy (\( Acc \)) score reflects the ability of a measurement to match the actual value of the quantity being measured (Metz, 1978). The overall accuracy score of the entire classification problem can be computed by macro-average (\( Acc_{\text{ME}} \)) as follows:

\[
Acc_g = \frac{TP_g + TN_g}{TP_g + FP_g + TN_g + FN_g}, \quad Acc_{\text{ME}} = \frac{\sum_{g \in G} Acc_g}{|G|}
\]

**Predicted results with a range of lessons**

The aim of this section is to focus on the change in final student grades predicted in each lesson to keep track of student’s learning situation and to grasp their characteristics in consecutive lessons. Through the prediction process, a teacher can understand students’ learning situations, attitudes and motivation and give correct assessment and useful feedback to them. We applied the majority vote (MV) method (Luo, Sorour, Goda, & Mine, 2015) to each student’s grades predicted in a range of lessons. We separated the prediction results into different lesson intervals, as shown in Figure 5.

We define the MV method as follows:

Let \( G \) be a set of grades \( \{ g_0, g_1, g_2, g_3, g_4 \} \); each element of \( G \) corresponds to each grade, i.e., \( g_0, g_1, g_2, g_3, \) and \( g_4 \) correspond to S, A, B, C and D, respectively.

Let \( MV_L(m,n) \) be the function of MV of student \( k \) from lessons \( m \) to \( n \). \( MV_L(m,n) \) returns a set of predicted grades for student \( k \) whose occurrence frequency from lessons \( m \) to \( n \) was the greatest.

**Definition 1.** \( MV_L(m,n) \)

\[
MV_L (m,n) = \{ g_i \mid \arg \max_{g_i \in G} f(k, g_i)(m,n) \}
\]

where \( f(k, g_i)(m,n) \) returns the occurrence frequency of predicted grade \( g_i \) of student \( k \) from lessons \( m \) to \( n \). For example, if the predicted grades of student 1 from lessons 1 to 3 are respectively S (= g_0), A (= g_1), B (= g_2), then \( f(1, g_0)(1, 3) = 1, f(1, g_1)(1, 3) = 1, \) and \( f(1, g_2)(1, 3) = 1 \). So, \( MV_L(1,3) \) returns \{ g_0, g_1, g_2 \}.

Function \( \delta \) returns a score according to the results returned by the MV function. \( \delta \) is defined in Definitions 2.

Here we use the notation \( |.| \) that denotes the cardinality of a set. For example, if \( MV_L(1,3) \) returns \{ g_0, g_1, g_2 \}, then \( |MV_L(1,3)|=3 \).

**Definition 2. \( \delta \)**

\( \delta \) (\( MV_L(m,n) \)) returns 1 if \( g_i \) is the actual grade of student \( k \), \( g_i \in MV_L(m,n) \), \( |MV_L(m,n)| \leq 2 \), and \( g_i \notin MV_L(m,n) \) such that \( |.-i| > 1 \), 0 otherwise. For example, we assume that the actual grade of student \( k \) is \( g_0 \), if \( MV_L(m,n) = \{ g_0, g_2 \} \), then \( \delta (MV_L(m,n)) = 1 \).
\{g_0, g_1\}, then \(\delta(MV_k(m,n)) = 1\). If \(MV_k(m,n) = \{g_0, g_1, g_2\}\) then \(\delta(MV_k(m,n)) = 0\). If \(MV_k(m,n) = \{g_0, g_2\}\) then \(\delta(MV_k(m,n)) = 0\).

Definition 3. \(TP(m,n)\)

We define \(TP(m,n)\) that returns the \(TP\) rate from lessons \(m\) to \(n\) as follows:

\[
TP(m,n) = \sum_{k=1}^{N_s} \delta(MV_k(m,n)) / N_s
\]

where \(N_s\) is the number of students. Next, we evaluate \(F_1^{Ma}\) in different interval groups of consecutive lessons.

In the current experiment, we dealt with students who had not submitted their comments as grade D for Classes A and B and grade C for Class C, because the behavior of not submitting comments can be regarded as most undesirable. It also ensures that for each lesson, every student has one predicted grade. We calculated the \(TP\) rate and \(F_1^{Ma}\) of the pLSA and the LDA models with three viewpoints: P, C and N, by changing the range of lessons between lesson 1 and 15.

Results

This section will report the results of final student grade prediction. The analysis was run in two steps. The first step conducts the comparison of the pLSA model with that of the LDA model in predicting final student grades. The second step examines the effect of the \(MV\) method with different lesson intervals.

Prediction performance in each lesson

Tables 5 and 6, and Figure 3 show the \(F_1^{Ma}\) and \(Acc^{Ma}\) results of final student grades for Class A, B and C with the pLSA and the LDA models. We can see that the LDA model had higher prediction results for all three sets of class comments than the pLSA model. C-comments achieved the best results among the three-viewpoints and the results with Class C outperformed those with Class A and B, where C-comments of Class C consist of the comments of Q2, Q3 and Q4.

<table>
<thead>
<tr>
<th></th>
<th>(F_1^{Ma})</th>
<th>(Acc^{Ma})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>73.7</td>
<td>71.2</td>
</tr>
<tr>
<td>Class B</td>
<td>69.1</td>
<td>67.7</td>
</tr>
<tr>
<td>Class C</td>
<td>73.5</td>
<td>75.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(F_1^{Ma})</th>
<th>(Acc^{Ma})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>85.6</td>
<td>81.5</td>
</tr>
<tr>
<td>Class B</td>
<td>79.3</td>
<td>78.9</td>
</tr>
<tr>
<td>Class C</td>
<td>86.3</td>
<td>85.0</td>
</tr>
</tbody>
</table>

The characteristics of data for each class

To show the differences in student grade prediction performance between the three classes, Table 7 presents the average overall prediction results with the LDA and SVM model. Class C achieved the highest results. Further, the \(F_1^{Ma}\) and \(Acc^{Ma}\) for Class A were higher than Class B. Also, we examined the prediction results after merging Class A and B to show the effect of all data. The prediction results were 84.3 \% for \(F_1^{Ma}\) and 81.2 \% for \(Acc^{Ma}\). Figure 4 displays the \(F_1\) results from lessons 1 to 15. As the results indicate, we cannot see the big difference between the higher and lower grades on their prediction performance. Also, the difficulty of the subject seemed
to influence the prediction performance, as the prediction results in lesson 7 for classes A and B dropped from those in lesson 6.

![Figure 3. Overall $F_1^{Ma}$ results with C-comments](image)

**Table 7. Overall prediction results using LDA model**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Class A</th>
<th>Class B</th>
<th>Class A &amp; B</th>
<th>Class C</th>
<th>Class A</th>
<th>Class B</th>
<th>Class A &amp; B</th>
<th>Class C</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>87.8</td>
<td>82.8</td>
<td>86.6</td>
<td>85.9</td>
<td>83.1</td>
<td>82.3</td>
<td>82.6</td>
<td>84.2</td>
</tr>
<tr>
<td>A</td>
<td>88.1</td>
<td>84.9</td>
<td>87.2</td>
<td>89.2</td>
<td>84.9</td>
<td>83.3</td>
<td>82.6</td>
<td>87.7</td>
</tr>
<tr>
<td>B</td>
<td>86.9</td>
<td>81.9</td>
<td>85.5</td>
<td>88.9</td>
<td>84.8</td>
<td>83.6</td>
<td>83.4</td>
<td>88.5</td>
</tr>
<tr>
<td>C</td>
<td>85.5</td>
<td>80.1</td>
<td>83.1</td>
<td>86.3</td>
<td>83.0</td>
<td>84.1</td>
<td>81.3</td>
<td>84.4</td>
</tr>
<tr>
<td>D</td>
<td>84.2</td>
<td>79.3</td>
<td>78.9</td>
<td></td>
<td>83.3</td>
<td>80.3</td>
<td>76.8</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 4. Overall $F_1$ results with LDA and C-comments](image)

**Predicted results after different lesson intervals**

We employed the MV method to get a continuous track of student performance. Here, we considered $TP$ rates and F-measure results and investigated the effect of different lesson intervals. Tables 8 and 9 show the average $TP$ rate and F-measure of student grades for the three classes with three viewpoints: P, C- and N using the pLSA
and the LDA models. The results indicate that the performance increased compared to the prediction results for each lesson, and LDA with C-comments had the highest results.

**Table 8. Correctly predicted TP rate (%) of student grade**

<table>
<thead>
<tr>
<th>Class</th>
<th>pLSA</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>C</td>
</tr>
<tr>
<td>Class A</td>
<td>80.0</td>
<td>83.3</td>
</tr>
<tr>
<td>Class B</td>
<td>76.2</td>
<td>81.0</td>
</tr>
<tr>
<td>Class C</td>
<td>82.0</td>
<td>86.5</td>
</tr>
</tbody>
</table>

**Table 9. F-measure results in consecutive lessons**

<table>
<thead>
<tr>
<th>Class</th>
<th>pLSA</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>C</td>
</tr>
<tr>
<td>Class A</td>
<td>78.4</td>
<td>79.6</td>
</tr>
<tr>
<td>Class B</td>
<td>74.9</td>
<td>79.2</td>
</tr>
<tr>
<td>Class C</td>
<td>79.2</td>
<td>84.4</td>
</tr>
</tbody>
</table>

**Figure 5. F-measure results for different interval groups of consecutive lessons**

Figure 5 shows $F_1^{MV}$ of the MV results for the three classes using LDA with C-Comments in L(1, s), which is between lessons 1 and s, where s varies from 3 to 15. Class C had the highest $F_1^{MV}$ rate, which reached to 93.4% in L(1-15), and the $F_1^{MV}$ were 88.5% and 87.5% for Class A and Class B, respectively. From Figure 5(a), we can see that $F_1^{MV}$ from L(1-3) was lower than that from L(1-5). This means that it was the first experience for the students to describe their learning situations, which affected their writing. As the lessons go by, the prediction results for the three Classes in the first 6 lessons became better. From lesson 7, the prediction results decreased for Class A and Class B due to the beginning of a new subject, which is on C programming and more difficult than previous lesson subjects. Figure 5(b) shows $F_1^{MV}$ in three consecutive lessons whose end lesson varies from 3 to 15. The $F_1^{MV}$ results for Class A and B were decreased in L(7, 9) from L(4, 6). We believe that the difficulty of the new subject affected the $F_1^{MV}$. Further, Figure 5(c) displays $F_1^{MV}$ by considering five lessons. The $F_1^{MV}$ reached 88.8%, 84.8% and 91.1% from L(11-15) for Classes A, B and C, respectively.
Discussion

Building a student performance prediction model has long been an important research topic in many academic disciplines. The present study uses different types of free-style comments, since the potentials of these comments can reflect student learning attitudes and activities connected with the lessons. In this section, we discuss the answers of the four research questions described in Introduction.

Question 1

The results shown in Tables 5, 6 and Figure 3 answer the research question that the LDA model had more significant results than the pLSA model in predicting final student grades for Classes A, B and C, and the current results consistent with (Sorour, Luo, Goda, & Mine, 2015). In addition, the prediction results of the C-comments were better than those of the P- and the N-comments. Students described the current activity more clearly, which more clearly indicated their grades; and this tendency was reflected in the prediction accuracy and F-measure of their grades.

Question 2

For the purpose of collecting high quality comment data that reflects students’ attitudes and activates, the PCN method was introduced through five questions used for Class C. The results indicated that all the prediction results for the three classes were high and the difference in results between the three classes was small, as shown in Table 7 and Figure 4; the PCN method can help to obtain good prediction performance even if classes, teachers, students, and formats of PCN items are different.

From the previous results, we can conclude that collecting good comment data is not an easy task. The instructor should give some pointers that lead students to describe their comments. Class C students described their learning situations using more specific topics (understanding, finding and collaboration). This kind of comment will help teachers to focus on the most important attributes for each group and to automatically build a dictionary that includes general and specific factors. Also, Class A outperformed Class B. Although students took the same subject given by the same teacher, they wrote different comments. Furthermore, we can distinguish upper grade students in Class A. They expressed their positive attitudes (able to – interesting – work hard – solve a problem – can be) than Class B. Class B students tended to write more about lesson topics.

Question 3

The results displayed in Figure 5 illustrate the effect of the MV method in continuously tracking student performance. We can conclude that it is possible to make an early prediction after five lessons with good prediction accuracy.

Question 4

Research question 4 concerns the relationship between the difficulty of a subject and student grade prediction results. From Figures 3, 4 and 5, we assumed that the difficulty of the subject had influenced student attitudes to expressing their behavior and sometimes did not allow students enough leeway to write meaningful comments; they wrote better comments in classes A and B while learning Computer Literacy from lessons 1 to 6 than while learning C-programming in lessons 7 to 15. We can assume that the worsening of prediction results from lesson 7 was due to the nature of the comments; students started coding and the comments included additional noise, i.e., programming/technical content, as was the case with Class C from the first five lessons. The results improved by the middle of the semester.

Conclusions and future work

This paper discussed a methodology which connects perspectives of LA and DM to understand students’ situations more deeply and to solve the problem of predicting students’ performance. In the current study, we examined different formats of free-style comments through multiple class data by multiple teachers to
holistically quantify student’s attitudes and activities in the learning environment. Although it is not an easy task to collect comment data from students, the experimental results showed that F-measure/accuracy can reveal some of the substantial differences between classes, lessons and students’ grades. In addition, students’ comments from Class C, which were informed by detailed pointers regarding the comment format, were able to generate the highest prediction performance. Also, the upper and lower grade groups were distinguished through the extracted topics. On the other hand, we can conclude that the MV method improved the overall prediction results. It enabled a continuous track of student performance and dealt with the prediction error accruing in each lesson. Finally, prediction results based on topic models have tightly different decisions. This finding helps an instructor to prepare his lesson and focus on the attributes of each group.

Further research should be conducted to address the limitations of this paper. First, it is necessary to understand the factors that affect students’ academic achievement (e.g., students’ attitudes “positive or negative,” their skills and efforts, which will help the teacher to understand the educational landscape and provide feedback after each lesson to improve students’ performance. In addition, we did not consider common attributes among the students that may be extracted from their comments. Second, it is difficult for teachers to interpret the current prediction models. We should build a student performance prediction model that is both practical and understandable. One direction for our future work is to extract some attributes from all students’ comments and generate a set of rules to predict students’ performance and give automated feedback so as to enhance awareness of learning for a greater number of students during the course.

Acknowledgements

This work was supported in party by JSPS KAKENHI Grant No. 26540183, 26350357 and 16H02926.

References


Minami, T., & Ohura, Y. (2013). Lecture data analysis towards to know how the students’ attitudes affect to their evaluations. In Proceedings of the 8th International Conference on Information Technology and Applications (pp. 164-169). Sydney, Australia: ICTIA.


Effects of the Team Competition-Based Ubiquitous Gaming Approach on Students’ Interactive Patterns, Collective Efficacy and Awareness of Collaboration and Communication

Chih-Hung Chen¹ and Gwo-Jen Hwang²

¹Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taipei, Taiwan // ²Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan // chihhung.chen@livemail.tw // gjhwang.academic@gmail.com

(Submitted October 23, 2015; Revised January 29, 2016; Accepted March 2, 2016)

ABSTRACT

Previous research has illustrated the importance of acquiring knowledge from authentic contexts; however, without full engagement, students’ learning performance might not be as good as expected. In this study, a Team Competition-based Ubiquitous Gaming approach was proposed for improving students’ learning effectiveness in authentic learning activities. An experiment on a natural science course was carried out in an elementary school to evaluate the effects of the approach on students’ awareness of collaboration, communication and collective efficacy. Moreover, the students’ learning behavioral patterns were explored by analyzing their interactive logs in the discussion forum. The experimental results indicated that, in addition to learning achievements, the proposed approach significantly enhanced the students’ collective efficacy as well as their awareness of collaboration and communication; moreover, the students who learned with the proposed approach showed more effective behavioral patterns than those who learned with conventional team-based ubiquitous learning.

Keywords

Competition game, Team-based learning, Mobile learning, Ubiquitous learning, Collective efficacy

Introduction

In the past decade, the potential and effectiveness of using handheld devices with emerging facilities such as wireless communication and sensing technologies to support students’ learning in authentic contexts has been reported by researchers (Chiang, Yang, & Hwang, 2014; Ozdamli & Uzunboylu, 2015). Via these technologies, students can obtain instant assistance or prompts and access digital materials when learning in related real-world contexts. Such a learning approach has been called context-aware ubiquitous learning (Chen, Chang, & Wang, 2008; Hwang, Tsai, & Yang, 2008). Although the importance of situating students in authentic contexts with support from the learning system has been emphasized, researchers have stated the challenges of learning in such complex learning scenarios consisting of real-world and digital-world contexts and learning resources (Chu, 2014; Kearney, Burden, & Rai, 2015). One of the challenges is the difficulty of having students pay full attention to the ubiquitous learning (u-learning) tasks since there could be many messages or interruptions from the environment they are situated in (Chen, Hwang, & Tsai, 2014). Another challenge is to help students deal with frustrations during the learning process. Thus, it is important to develop a u-learning approach that can encourage students to fully engage in the learning tasks and help them deal with the problems encountered during the learning process (Hsu & Hwang, 2014).

Among various learning support approaches, collaborative learning is considered an effective strategy for assisting students’ learning in real-world environments. A well-designed collaborative learning activity can contribute to students’ learning effectiveness and encourage them to become active problem solvers (de Laat & Lally, 2003); moreover, collaborative learning activities create more opportunities for individuals to share information, negotiate with peers and make decisions (Curtis & Lawson, 2001). Nevertheless, simply situating students in team-based learning tasks does not necessarily achieve collaboration effects (Gelmini-Hornsby, Ainsworth, & O’Malley, 2011). Researchers have indicated that it is crucial to consider the quality of students’ social interactions as well as their engagement when developing collaborative learning activities (Khosa & Volet, 2014; Marcos-García, Martínez-Monés, & Dimitriadis, 2015).

Digital game-based learning (DGBL) is an effective educational approach which is comprised of some enjoyable elements for engaging students in challenging activities (Prensky, 2007). Scholars have pointed out that DGBL, which is regarded as a kind of potential learning approach for promoting students’ learning interest and motivation, may provide opportunities for enhancing students’ social interactions and engagement in collaborative learning (Ardito, Lanzilotti, Costabile, & Desolda, 2013; Brom et al., 2014; Sung & Hwang, 2013).
The awareness of communication, collaboration and collective efficacy are very important to students since it is the basis of fostering communication and collaboration abilities. Therefore, in this study, a team competition-based ubiquitous gaming approach (TCUG) is proposed to engage students in active learning behaviors and collective effective efficacy, as well as improving their awareness of collaboration and communication.

Literature review

Computer-Supported Collaborative Learning

Collaborative learning, a form of interaction among students, refers to the methodologies designed to engage students in collaboratively solving problems (Eryilmaz et al., 2013; Lai & Hwang, 2014; So & Brush, 2008). Collaborative learning, if appropriately conducted, can not only contribute to students’ knowledge sharing and transfer (Noroozi, Teasley, Biemans, Weinberger, & Mulder, 2013), but can also promote their higher order abilities (Chen, Wang, & Lin, 2015). That is, collaborative learning could benefit students by promoting their learning outcomes, collaboration and communication ability (Reeves, Herrington, & Oliver, 2004).

One important factor of collaboration is the discussion that occurs during task engagement (Curtis & Lawson, 2001), which facilitates the sharing of information, negotiation with peers, and decision making (de Laat & Lally, 2003). Gelmini-Hornsby, Ainsworth and O’Malley (2011) described that students can benefit from collaboration when they engage in an effective discussion. Proper instructional methods engage students in socially interacting with peers for questioning as well as elicitation (Remesal & Colomina, 2013; Wang & Hwang, 2012).

One problem of collaborative learning is that students’ active interactions with team members are generally rather poor (Crook, 1998). Moreover, lack of engagement in collaborative tasks could result in freeriding or social loafing, that is, the phenomenon of reducing individual striving when working collectively compared with when individually acting on the same learning activity (Gu, Shao, Guo, & Lim, 2015; Triantafyllakos, Palaigeorgiou, & Tsoukalas, 2011).

Although Computer-Supported Collaborative Learning (CSCL) systems seem to be promising in terms of facilitating peer interactions, the result of a CSCL activity could be disappointing without a careful learning design (Gelmini-Hornsby, Ainsworth, & O’Malley, 2011; Slof, Erkens, Kirschner, Janssen, & Phielix, 2010). For example, Wang (2010) provided students with an online shared workspace for collaboration, and asserted that only about half of the groups actively shared resources and negotiated ideas via the workspaces. Therefore, several scholars have indicated the necessity of incorporating effective learning strategies or tools into CSCL activities.

Digital game-based learning

Digital game-based learning (DGBL) refers to those learning activities which integrate computer games into a learning environment (Akcaoglu, 2014; Hwang, Chiu, & Chen, 2015; Prensky, 2007). It has the potential advantages of affording students interactive learning opportunities for stimulating their collaboration (Sung & Hwang, 2013) by situating them in a scenario-based task so as to encourage their active learning and motivate their engagement and satisfaction (Hwang & Chang, 2015; Yang & Chang, 2013). For example, Yang (2015) asserted that digital games provide plentiful opportunities for students’ collaboration and sharing of experiences via linking learning with authentic duties, while Li and Tsai (2013) described that digital games can stimulate collaboration due to their characteristics such as instant feedback, clear goals and rules.

In recent years, several studies have taken advantage of digital games in educational environments with the aim of improving students’ learning effectiveness. For example, Hwang, Wu and Chen (2012) developed a competitive online board game to promote not only students’ learning motivation but also their learning achievement. Ardito, Lanzilotti, Costabile and Desolda (2013) integrated educational games with the discovery learning strategy to help consolidate knowledge stimulating students’ collaboration.

Although previous research has described the potential of DGBL for promoting students’ learning motivation and fostering their collaborative learning skills (e.g., Sung & Hwang, 2013), there have been few studies on the effects of integrating DGBL and CSCL in the context-aware ubiquitous learning environment. Some previous studies have reported the behavioral analysis results of in-field learning (e.g., Chiang, Yang, & Hwang, 2014);
however, to the best of our knowledge, none of them have investigated students’ learning behaviors in collaborative game-based in-field learning, and neither have they evaluated students’ perceptions of collaboration and communication. Thus, in this study, a TCUG approach is proposed to enhance elementary school students’ collective efficacy as well as their awareness of collaboration and communication, which are likely to affect their corresponding competences in the future, as well as their learning behaviors and performance on field trips.

**Research questions**

An experiment was conducted to evaluate the effects of the TCUG approach via answering the following research questions.

- Do the students using the TCUG approach display better awareness of collaboration than those using conventional collaborative u-learning?
- Do the students using the TCUG approach display better awareness of communication than those using conventional collaborative u-learning?
- Do the students using the TCUG approach have higher collective efficacy than those who learn with conventional collaborative u-learning?
- Do the students using the TCUG approach show better learning achievement than those learning with conventional u-learning?
- What are the differences between the interactive patterns of the students using the TCUG approach and those using conventional collaborative u-learning?

**Context-aware collaborative learning system with a team competition-based gaming approach**

Based on the TCUG approach, a team competition-based ubiquitous gaming system has been developed utilizing Visual Studio C# and Android Studio. QR codes are used to guide students to observe specific ecological subjects or phenomena in the field. Each student is provided with a tablet computer to interact with the real-world environment, while a wireless network is provided to not only facilitate their communication with their peers via an online forum but also to enable communication between the tablet computers and the server.

The learning materials of “biology and environmental science” are integrated into gaming scenarios which guide students to explore and observe in a real-world environment on an elementary school campus. The gaming story is related to a southern island country, called “Biodiversity Island.” The king of the country, whose name is Formosa, has magic power, so the island is not only full of life and vitality but is also a diverse biological environment. However, the shelters are disappearing due to the excessive use of energy, and the country has been in a state of disarray. A witch, Katrina, takes advantage of the disarray and smuggles in a number of invasive species in order to cause serious destruction to the ecological balance. Therefore, the king dispatches some brave residents, the role students take, to find the reasons for the environmental changes in the ecological systems and to propose solutions to the problems. Students must complete seven tasks in order to recover the king’s magic power and restore the vitality of “Biodiversity Island.”

A total of 7 gaming tasks were designed for this game, namely “Alien Invasion,” “Dark Forces,” “Hostile Water,” “Dark Breakages,” “Airspace Battle,” “Territory Battle” and “Honor Battle.” Each time the player is allowed to select one of the gaming tasks, except for the last one called “Honor Battle.” After the team chooses a gaming task, such as “Alien Invasion,” the compelling storyline of the gaming task is shown: “A number of invasive species are smuggled in and hence cause serious destruction to the ecological balance.”

Following that, a gaming task consisting of an opening question of the problem-solving activity is displayed: “Lantana camara is considered a kind of invasive species; please explore at least five characteristics of Lantana camara.” The team is then guided to find the related learning target in the real-world environment via the map provided by the learning system. When students arrive at the location of the learning target for observation, they are asked to scan a QR-code tag on the target to confirm their physical location, as shown in Figure 1.

Figure 2 shows the main gaming interface, which contains the areas of task description, task prompt, chat room, team ranking, gameplay progression, help seeking, timing and scoring. The members in a team need to collaborate to collect the required data for the gaming tasks in order to get high scores; moreover, a team can only submit an answer if all of the team members agree with the answer. During the gaming activity, the students
can discuss face to face or via an online chat room, which enables team members to discuss and share their findings instantly. In addition, to remind the students of the competitive status among the teams, the current top-five gaming scores of all teams are displayed on the gaming interface.

The learning team can collaboratively complete the task with the assistance of some tools or NPCs (Non-Player Characters) which are provided by the learning system. For example, a blue shield can provide students with the key characteristic of the creatures and environment, while a red shield prompts them to make comparisons between two kinds of creature or environment with opposite characteristics. A treasury of knowledge provides information about a creature or environment, and a sage can offer some key clues for observing a creature or environment. Moreover, a crystal ball provides the opportunity to find data by connecting to a web search engine.

“Lantana camara has been considered as a kind of the invasive species; please identify at least five characteristics of Lantana camara.”

Figure 1. A learning task corresponding to the learning park

Figure 2. The main Team Competition-based gaming system interface
The solutions to the task must be proposed within five minutes, and two tests have to be completed within a further five-minute period. One is a basic test of recognition and the other is a comparative test related to two creatures or environments. For example, the learning system displays a basic test: “Which one is the key reason for the rapid multiplication of Lantana Camara?” All tests are multiple-choice questions, and the submission is effective after the team members reach a consensus regarding the answer. To facilitate the team members’ consultation with each other, the individual answer is shown on the forum. If a team gives a wrong answer, it is allowed to resubmit the answer. The earlier a team gives the right answer, the more points it gets. After a team successfully completes the task, a compelling storyline about winning this gaming task battle is presented.

After completing the first 6 gaming tasks, the team is allowed to challenge the last one that is the “Honor Battle.” Once a team completes this task related to solar energy and energy efficiency, the vitality of “Biodiversity Island” is restored and the game is completed.

Method

Participants

The participants in this study were four classes of sixth graders (11.5 years old on average) taking a natural science course for four periods a week in an elementary school in northern Taiwan. The four classes consisted of a total of 101 students who were taught by the same teacher who has more than ten years’ science teaching experience in the elementary school. Two classes were assigned to be the experimental group (n = 50) who learned with the team competition-based collaborative ubiquitous learning, while the other two were the control group (n = 51) who learned with the collaborative ubiquitous learning. During the learning activity, the students in both groups were assigned to small learning teams, each of which had 3 or 4 team members with different knowledge levels.

Experimental procedure

The collaborative ubiquitous learning activity about the “biology and environmental science” unit of the elementary school natural science course was conducted to evaluate the effects of the team competition-based gaming approach. The students collaboratively completed the tasks in a real-world learning environment via the guidance of the Team Competition-based Ubiquitous gaming system; moreover, a fantasy story based on the authentic environment and learning topic was displayed to engage the students in completing the learning tasks.

Before the experiment, the students were instructed via a regular eight-period course to learn the fundamental knowledge of biodiversity. Following that, they filled out the pre-questionnaire of collaboration, communication competences and collective efficacy.

Afterwards, the participants started to deal with the collaborative ubiquitous learning tasks. The experimental group used the team competition-based gaming approach. On the other hand, the control group learned with the collaborative ubiquitous learning approach without the gaming mechanism; that is, they were asked to complete the same learning tasks in the same real-world learning environment. In addition, the learning materials and guidance or hints provided to both groups were identical. To engage the students in deep research on the creatures and the learning environment, all teams were subsequently required to complete a report about environmental changes during the activity.

After the learning activity, the students filled out the post-questionnaires of collaboration, communication competences and collective efficacy. Finally, nine students were randomly selected from each group for a one-on-one interview to collect their learning perceptions of and opinions on the learning activity.

Measuring tools

The measuring tools adopted in this study included the questionnaires of collaboration, communication competences, collective efficacy, pre-test, post-test, and interactive content coding scheme.

The questionnaire of students’ awareness of their collaboration was designed based on the study announced by Lai and Hwang (2014). It consists of five items with a five-point Likert rating scale, such as “When I worked...”
with my team members, I think our conversation was good” and “When peers propose their opinions, I will not question their motivation.” The Cronbach’s alpha value of the questionnaire stated by the original study was 0.85, implying highly acceptable reliability in internal consistency.

The survey of students’ awareness of their communication was adopted based on modifying the measurement developed by Lai and Hwang (2014). A total of five items (e.g., “When I talk with someone, I feel comfortable.” and “When I talk with peers, I will consider their feelings.”) made up the questionnaire with a five-point Likert scale. The Cronbach’s alpha coefficient stated by the original study was 0.88, implying highly acceptable reliability of the students’ communication questionnaire.

To investigate students’ collective efficacy, the survey developed by Wang and Hwang (2012) was adopted, which was based on the questionnaire proposed by Pintrich, Smith, Garcia and McKeachie (1991). This survey consists of eight items with a five-point Likert scale, such as “I am confident that our team can understand the most complicated part of this work through teamwork” and “I am confident that our team can finish this work well.” The alpha reliability score described in the original study was 0.92, indicating highly acceptable reliability in internal consistency.

To evaluate students’ learning achievement, the pre-test and post-test were designed by two natural science teachers with years of teaching experience. The pre-test, which comprised 20 multiple-choice items, was related to the fundamental knowledge of biodiversity. On the other hand, the post-test, which consisted of 26 multiple-choice items and four short answer questions, was about the course of “creatures and environment.” The perfect score of each test was 100.

Regarding the analysis of students’ interactive content, this study utilized the Interaction Analysis Model (IAM) coding scheme, developed by Gunawardena, Lowe and Anderson (1997). This scheme, which has been adopted in some of the previous research (Chiang, Yang, & Hwang, 2014; Hou, Chang, & Sung, 2008; Lan, Tsai, Yang, & Hung, 2012) was used to analyze students’ cognitive behavior on the forum. IAM divided the discussion content into five levels, as shown in Table 1.

<table>
<thead>
<tr>
<th>Code</th>
<th>Level</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0</td>
<td>Content irrelevant to the learning topic</td>
<td>Content that is completely unrelated to what is being discussed or considered.</td>
</tr>
<tr>
<td>L1</td>
<td>Sharing or comparing information</td>
<td>Simply explaining the observation or opinion; Agreeing with a peer’s idea.</td>
</tr>
<tr>
<td>L2</td>
<td>Exploration of inconsistency among team members</td>
<td>Illustrating the dissonance between different ideas or opinions; Questioning about the topic and proposing one’s opinions.</td>
</tr>
<tr>
<td>L3</td>
<td>Negotiation or construction of knowledge</td>
<td>Abstracting and illustrating the issues or opinion on the topic.</td>
</tr>
<tr>
<td>L4</td>
<td>Comparison and modification of the proposed synthesis or newly constructed knowledge</td>
<td>Comparing the conclusions of the discussions with existing knowledge; Proposing individual points in contrast with personal experience or references.</td>
</tr>
<tr>
<td>L5</td>
<td>Agreement or application of newly constructed meaning</td>
<td>Inducing the discussions; Proposing deeper reflection, new construction or metacognition.</td>
</tr>
</tbody>
</table>

### Experimental results

#### Awareness of collaboration

A one-way ANCOVA was utilized to measure the students’ awareness of collaboration competence using individual pre-questionnaire scores as a covariate, ubiquitous learning approaches as an independent variable, and the post-questionnaire score as a dependent variable.

After verifying that the assumption of homogeneity of regression was not violated with $F = 0.09$ ($p > .05$), the post-questionnaire scores were analyzed with ANCOVA. As shown in Table 2, a significant impact of the ubiquitous learning approach was found on the two groups with $F = 6.21$ ($p < .05$, $\eta^2 = 0.060$). Moreover, the adjusted mean values and standard deviation errors of the students’ perception scores were 4.12 and 0.10 for the experimental group, compared with 3.78 and 0.10 for the control group. It was confirmed that the awareness of
collaboration of the students using the team competition-based gaming approach was significantly higher than that of the students adopting the collaborative ubiquitous learning approach. Furthermore, according to the definition proposed by Cohen (1988), the ANCOVA results of the ubiquitous learning model gave a moderate effect size with $\eta^2 > 0.059$.

**Table 2. The ANCOVA results for the students’ awareness of collaboration**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>50</td>
<td>4.12</td>
<td>0.69</td>
<td>4.12</td>
<td>0.10</td>
<td>6.21*</td>
<td>0.060</td>
</tr>
<tr>
<td>Control</td>
<td>51</td>
<td>3.78</td>
<td>0.86</td>
<td>3.78</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. $p < .05$.

**Awareness of communication**

As for the impact of the Team Competition-based Ubiquitous gaming system on the students’ awareness of their own communication competence, a one-way ANCOVA was conducted by adopting ubiquitous learning approaches as an independent variable, while the post-questionnaire scores were a dependent variable and the pre-questionnaire scores were a covariate.

In order to judge whether the use of ANCOVA was proper, the homogeneity test was computed first. It was found that the assumption of homogeneity of regression was passed with $F = 0.43$ ($p > .05$). Following that, ANCOVA was performed and a significant difference ($F = 8.73, p < .01, \eta^2 = 0.082$) was proved between the two groups (as shown in Table 3). The adjusted mean values of the students’ awareness of communication ratings for the experimental group and control group were 4.35 (Std. error = 0.12) and 3.86 (Std. error = 0.12), respectively. It was found that the Team Competition-based Ubiquitous gaming system had a significantly positive effect on the students’ awareness of their communication competence in the real-world ubiquitous learning activity.

**Table 3. The ANCOVA results for the students’ awareness of communication**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>50</td>
<td>4.36</td>
<td>1.00</td>
<td>4.35</td>
<td>0.12</td>
<td>8.73**</td>
<td>0.082</td>
</tr>
<tr>
<td>Control</td>
<td>51</td>
<td>3.85</td>
<td>0.85</td>
<td>3.86</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. **$p < .01$. 

**Collective efficacy**

The homogeneity of the regression slopes was confirmed with $F = 0.28$ ($p > .05$), and afterwards ANCOVA was performed. The ANCOVA results in Table 4 describe the significant difference between the two groups ($F = 10.36, p < .01, \eta^2 = 0.096$) with moderate effect size; moreover, the adjusted mean values and standard deviation errors of the students’ collective efficacy ratings were 4.24 and 0.08 for the experimental group, and 3.86 and 0.08 for the control group. Accordingly, the collective efficacy of the students using the team competition-based gaming approach was significantly higher than that of the students using collaborative ubiquitous learning. In other words, the Team Competition-based Ubiquitous gaming system can effectively promote students’ collective efficacy in a real-world ubiquitous learning environment.

**Table 4. The results of the ANCOVA on students’ collective efficacy**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>50</td>
<td>4.21</td>
<td>0.64</td>
<td>4.24</td>
<td>0.08</td>
<td>10.36**</td>
<td>0.096</td>
</tr>
<tr>
<td>Control</td>
<td>51</td>
<td>3.88</td>
<td>0.77</td>
<td>3.86</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. **$p < .01$. 

**Learning achievements**

To analyze the influence of the TCUG approach on the learning achievement of the students with different pre-test scores, students in each group were divided into two subgroups, namely “high-pre-test” and “low-pre-test,” based on their pre-test scores. Furthermore, the one-way ANCOVA was utilized by taking the pre-test scores as a covariate, while the ubiquitous learning approaches were an independent variable and the post-test scores were a
dependent variable. The assumptions of homogeneity of regression were assessed for the two levels, and the $F$ values for “high-pre-test students” level and “low-pre-test students” level were 0.06 ($p > .05$) and 0.26 ($p > .05$), respectively. Accordingly, it was reasonable to execute ANCOVAs to evaluate the different pre-test levels of students’ learning achievement for the TCUG approach.

As described in Table 5, the significant effects of the ubiquitous learning approaches were found to be $F = 4.18$ ($p < .05$; $\eta^2 = 0.077$) for the high-pre-test students and $F = 11.76$ ($p < .01$; $\eta^2 = 0.207$) for the low-pre-test students, indicating that the post-test scores of these two pre-test levels were significantly different. Moreover, the adjusted means of the high-pre-test level’s post-test scores were 87.86 (Std. error = 2.17) for the experimental group and 81.22 (Std. error = 2.21) for the control group, indicating that the TCUG approach can improve high-pre-test students’ learning achievement. On the other hand, the low-pre-test students (adjusted mean = 82.54, Std. error = 2.21) who participated with the TCUG approach outperformed those who participated with collaborative ubiquitous learning (adjusted mean = 71.74, Std. error = 2.12) in terms of their learning achievement. According to Cohen’s (1988) declaration, the ANCOVA results of the TCUG approach represented a moderate effect size ($\eta^2 > 0.059$) and large effect size ($\eta^2 > 0.138$) for the high-pre-test level and low-pre-test level students, respectively.

### Table 5. The results of the ANCOVAs on students’ learning achievement

<table>
<thead>
<tr>
<th>Level</th>
<th>Group</th>
<th>$N$</th>
<th>Mean</th>
<th>$SD$</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>$F$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-pre-test level</td>
<td>Experimental group</td>
<td>27</td>
<td>86.78</td>
<td>9.15</td>
<td>87.86</td>
<td>2.17</td>
<td>4.18*</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>26</td>
<td>82.35</td>
<td>12.38</td>
<td>81.22</td>
<td>2.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-pre-test level</td>
<td>Experimental group</td>
<td>23</td>
<td>79.91</td>
<td>10.11</td>
<td>82.54</td>
<td>2.21</td>
<td>11.76**</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td>Control group</td>
<td>25</td>
<td>74.16</td>
<td>14.27</td>
<td>71.74</td>
<td>2.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $p < .05$; **$p < .01$.

### Interactive behaviors in the online forum

A total of 585 messages (experimental group 313, control group 272) in the chat room were recorded in the learning portfolio database, all of which were coded based on the IAM scheme. It was found that 94.57% and 95.99% of the peer interactions in the experimental group and control group, respectively, were scored as L1 or L2; that is, most of the students spent much time on lower-level social knowledge construction behaviors. The inter-rater kappa reliability of the IAM coding for the experimental group and the control group was 0.77 and 0.75, respectively. Accordingly, excellent inter-rater reliability ($k > 0.75$) and fair to good reliability ($0.40 < k < 0.75$) were confirmed based on the description given by Fleiss, Levin and Paik (1981).

A lag sequential analysis was utilized to explore the participants’ behavioral patterns on the online forum. The adjusted residuals of the experimental group are illustrated in Table 6, and five sequences are confirmed to be statistically significant, including L1 $\rightarrow$ L1, L2 $\rightarrow$ L2, L2 $\rightarrow$ L4, L3 $\rightarrow$ L3 and L4 $\rightarrow$ L5.

### Table 6. Adjusted residuals of the experimental group for IAM

<table>
<thead>
<tr>
<th></th>
<th>L0</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>L5</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>L1</td>
<td>0.48</td>
<td>6.57*</td>
<td>-5.89</td>
<td>-1.53</td>
<td>-2.14</td>
<td>-1.14</td>
</tr>
<tr>
<td>L2</td>
<td>-0.40</td>
<td>-6.47</td>
<td>6.62*</td>
<td>0.40</td>
<td>2.64*</td>
<td>-0.57</td>
</tr>
<tr>
<td>L3</td>
<td>-0.20</td>
<td>-1.38</td>
<td>0.37</td>
<td>2.60*</td>
<td>-0.34</td>
<td>-0.28</td>
</tr>
<tr>
<td>L4</td>
<td>-0.10</td>
<td>-0.58</td>
<td>-0.71</td>
<td>-0.34</td>
<td>-0.18</td>
<td>6.98*</td>
</tr>
<tr>
<td>L5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note.* $p < .05$.

Based on the significant sequences, a behavioral transfer diagram of the experimental group during the online forum is schematized in Figure 3. The continuously repeated relationships are shown for social knowledge construction; for example, L1 $\rightarrow$ L1, L2 $\rightarrow$ L2 as well as L3 $\rightarrow$ L3. Continuous actions represented in L1, L2 and L3 indicated that the experimental group preferred completing one gaming task before conducting a further one. For instance, L1 to L1 (sharing or comparing information) described that the participants collaboratively searched for and shared the information about the topics for a while; peers illustrated the dissonance between different ideas or opinions, asking questions about the topic or proposing opinions, represented by the L2 to L2 sequence; moreover, they engaged in abstracting the issues of the topic, described by the L3 to L3 sequence. Moreover, the sequence L2 $\rightarrow$ L4 implied that individuals may propose their point in contrast to personal
experience or references via the exploration of peers’ inconsistencies. Following that, the agreement or application of newly constructed meaning may take place according to the L4 to L5 sequence.

The interaction behavioral patterns of the control group were also analyzed. Table 7 illustrates the adjusted residuals with three statistically significant sequences: L1 → L1, L2 → L2 and L0 → L0.

A behavior transfer diagram was depicted on the basis of the significant sequences, as shown in Figure 4. The continuously repeated relationships, which are similar to those of the experimental group, were confirmed via L1 → L1 and L2 → L2. However, there was no significant relationship from L3 to L3 (Negotiation or construction of knowledge). Instead, peers may continuously talk about something irrelevant to the learning topic, as revealed by the L0 to L0 sequence.

**Table 7. Adjusted residuals of the control group for IAM**

<table>
<thead>
<tr>
<th></th>
<th>L0</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>L5</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0</td>
<td>2.96*</td>
<td>-0.01</td>
<td>-0.97</td>
<td>-0.35</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>L1</td>
<td>-0.08</td>
<td>3.42*</td>
<td>-3.34</td>
<td>-0.94</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>L2</td>
<td>-0.94</td>
<td>-3.74</td>
<td>3.94*</td>
<td>1.30</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>L3</td>
<td>-0.32</td>
<td>-0.01</td>
<td>0.28</td>
<td>-0.35</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>L4</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>L5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note. *p < .05.

**Discussion and conclusions**

In this study, a team competition-based learning system was developed for assisting students’ learning in a real-world environment. An experiment was conducted on a natural science course in an elementary school. The experimental results indicated that, in addition to learning achievements, the TCUG approach significantly improved students’ awareness of their collaboration and communication competences as well as their collective efficacy. It was also found that the TCUG approach more significantly benefitted the low pre-test students’ learning achievement more than that of the high pre-test students, showing the positive effects of integrating the team competition-based gaming approach into ubiquitous learning activities on low-achievement students. Moreover, the students using the TCUG system apparently revealed more higher cognitive level interactive patterns than those who used the conventional u-learning system, such as from L2 “sharing or following information” to L4 “comparing and modifying newly constructed knowledge or proposing new perspectives,” and from L4 to L5 “inferring or making reflections based on newly constructed meanings.” In addition, the
repeated interactive behaviors L3 → L3 (i.e., Negotiation or construction of knowledge) also showed that the students in the experimental groups had better communication and collaboration than the control group. This implies that the TCUG approach can promote the quality of social interaction among peers in the chat room owing to the integration of the learning contexts and the gaming storyline.

Scholars have stated that it is important to effectively promote students’ higher cognitive-level interactions via appropriate assistance in a collaborative learning environment (Gelmini-Hornsby, Ainsworth, & O’Malley, 2011). By way of illustration, Strijbos (2011) claimed that participants’ engagement may affect the quality of the collaboration. From the behavior patterns of IAM in this study, it was found that the Team Competition-based Ubiquitous gaming system can promote peers’ social interaction, which is considered as an important factor in the effects on collaborative learning. This may be the reason why the TCUG system promoted the students’ awareness of the collaboration and communication. This is consistent with the finding asserted by Sánchez and Olivares (2011), who inferred that students’ awareness of their own collaboration performance might be promoted via a game-based mobile learning environment.

On the other hand, the behavior patterns also illustrate why the students in the TCUG group demonstrated higher collective efficacy than those in the collaborative ubiquitous learning group. Such a finding provides strong evidence to support those collective efficacy studies inferring that the quality of group interactions plays an important role in constructing and promoting collective efficacy in a computer-supported collaborative learning environment (Wang & Hwang, 2012; Wang, Hsu, Lin, & Hwang, 2014). Bandura (2000) indicated that perceived collective efficacy could act as a stimulant to groups’ motivational commitment to their missions and performance accomplishments. This could also be a reason why the TCUG approach facilitated the students’ collaboration performances as well as their learning achievements. Such a finding also confirms the finding of Wang and Hwang (2012), who indicated that collective efficacy significantly motivated group members’ commitment to collaborative tasks while also improving their collaborative performance.

As it is a challenging issue to integrate gaming scenarios into an authentic learning environment in order to enhance students’ learning effectiveness (Chen, Liu, & Hwang, 2015; Schmitz, Klemke, & Specht, 2014), the findings of this study provide a good reference for those who intend to develop effective and motivational learning environments for authentic activities. Moreover, the proposed approach could be advantageous to teachers who plan to situate students in collaborative in-field learning. In the meantime, from the learning behavior analysis, it was found that the participants spent much time on lower-level social knowledge construction behaviors in the online forum. Thus, it is worth exploring additional strategies to promote the quality of the peer discussion. For example, we are planning a follow-up study that aims to provide students with visualized knowledge construction tools to promote their behavioral levels for social knowledge construction in the discussion forum.

Acknowledgements

This study is supported in part by the Ministry of Science and Technology of the Republic of China under contract numbers NSC 102-2511-S-011-007-MY3 and MOST 104-2511-S-011-001-MY2.

References


Conversational Agents Improve Peer Learning through Building on Prior Knowledge

Stergios Tegos* and Stavros Demetriadis
School of Informatics, Aristotle University of Thessaloniki, Greece // stegos@csd.auth.gr // sdemetri@csd.auth.gr

*Corresponding author

(Submitted October 29, 2015; Revised March 23, 2016; Accepted April 8, 2016)

ABSTRACT

Research in computer-supported collaborative learning has indicated that conversational agents can be pedagogically beneficial when used to scaffold students’ online discussions. In this study, we investigate the impact of an agile conversational agent that triggers student dialogue by making interventions based on the academically productive talk framework. An experimental activity in the context of a university course involved 72 undergraduate students who discussed online in dyads. Two conditions were compared: (a) dyads who received a gentle interventions while working on a learning task (treatment) and (b) dyads who worked on the same task without any agent interference (control). Utilizing a concept map created by the course instructor, the conversational agent delivered unsolicited interventions that encouraged treatment students to build on their prior knowledge, linking their current contributions to the main domain principles discussed during the course. Study findings indicate that the agent intervention mode substantially improved both individual and group learning outcomes. Evidence suggests that the agent effect on learning performance was mediated by students’ explicit reasoning, which was also found to be enhanced in the treatment condition.

Keywords
Conversational agent, Computer-supported collaborative learning, Academically productive talk

Introduction

Conversational agents

In the field of technology-enhanced learning, pedagogical agents have been developed to serve a wide variety of instructional roles, such as expert, motivator, or mentor (Baylor & Kim, 2005). Conversational agents are typically regarded as a subgroup of pedagogical agents involving learners in natural language interactions (Kerly, Ellis, & Bull, 2009). Research has shown using conversational agents to engage learners in one-to-one (student-agent) tutorial dialogues to improve students’ comprehension and foster students’ engagement and motivation (Veletsianos & Russell, 2014).

During the past decade, researchers also focused on developing conversational agents for collaborative learning support (e.g., Kumar & Rosé, 2011). Despite the established cognitive and social benefits of computer-supported collaborative learning (CSCL), collaborative knowledge construction is not a given but depends on the quality of interactions taking place among learners (Dillenbourg & Tchounikine, 2007; Kreijns, Kirschner, & Jochems, 2002). Under this prism, well-targeted supportive interventions can be used as a method to increase the probability of constructive peer interactions occurring by means of stimulating cognitive processes, such as conflict resolution, mutual regulation or explicit explanation (Tchounikine, Rummel, & McLaren, 2010). Evidence suggests that conversational agents with social interaction capabilities can enhance learning and idea generation productivity by providing dynamic support for learners working together (Kumar & Rose, 2011; Kumar, Beuth, & Rosé, 2011). Chaudhuri et al. (2008) reveal that agents guiding peers through prescribed lines of reasoning on specific topics can improve learning performance. A study by Walker, Rummel, and Koedinger (2011) indicates that an agent displaying reflective prompts in a scripted peer tutoring activity can help students produce conceptually richer statements.

Academically productive talk

Another research direction has recently emerged focusing on an agile form of conversational agent support, which emphasizes the key role of social interaction in student engagement and learning (e.g., Adamson, Dyke, Jang, & Rosé, 2014). This approach draws on the academically productive talk (APT) framework, itself originating from a substantial body of work on useful classroom discussion practices and norms (Michaels, O’Connor, & Resnick, 2008). According to APT, a peer dialogue in class should be accountable to the learning
community, accurate knowledge and rigorous thinking, irrespective of the subject area (Sohmer, Michaels, O’Connor, & Resnick, 2009). In view of the above, peers should paraphrase and expand on each other’s ideas (i.e., being accountable to the learning community), support the validity of their claims making explicit references to a pool of knowledge accessible to the group (i.e., being accountable to accurate knowledge), and logically connect their statements through rigorous argumentation (i.e., being accountable to rigorous thinking).

Unlike other well-known discourse frameworks such as the IRE (Initiation, Response and Evaluation), the APT framework does not entail closing down a conversation after successfully eliciting a correct learner’s response; instead, APT aims to promote and scaffold open-ended discussions where learners explicate their reasoning, compare their contributions with their partners’ and construct logical arguments based on accurate evidence (Michaels, O’Connor, Hall, & Resnick 2010). Indeed, APT does not expect the teacher to maintain full control over learners’ discussions, and prioritizes reasoning over correctness. The importance of the explicit articulation of reasoning is universally acknowledged by researchers, despite the different conceptualization of studies exploring the key features of a productive peer dialogue (for example, “transactivity,” “group cognition,” and “productive agency”) (Stahl & Rosé, 2011). The explicitness of students’ reasoning has been shown to enhance conceptual knowledge acquisition and improve collaboration practices (Papadopoulos, Demetriadis, & Weinberger, 2013).

Nevertheless, promoting students’ reasoned participation and orchestrating academically productive discussions are often challenging tasks, requiring teachers to provide dynamic support via facilitative conversational moves (Sohmer et al., 2009). Table 1 presents a selection of those APT moves, which can be regarded as conversational interventions (or actions) aiming to model and trigger appropriate forms of peer dialogue (Michaels et al. 2010). Considering that such moves have proven conducive to scaffolding open-ended discussions in class, researchers further proceed to explore their potential contribution to CSCL by tailoring agent supportive mechanisms to display APT-based prompts (Dyke, Adamson, Howley, & Rosé, 2013).

### Table 1. A list of academically productive talk moves

<table>
<thead>
<tr>
<th>APT Move</th>
<th>Example</th>
<th>Accountability dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Link Contributions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Agree-Disagree</td>
<td>“Do you agree with what your partner said about …?”</td>
<td>Learning Community</td>
</tr>
<tr>
<td>B. Add-On</td>
<td>“Would you like to add something to …?”</td>
<td>Learning Community</td>
</tr>
<tr>
<td>2. Revoice</td>
<td>“So, are you saying that … Is that correct?”</td>
<td>Learning Community</td>
</tr>
<tr>
<td>3. Build on Prior Knowledge</td>
<td>“How does this connect with what we know about …?”</td>
<td>Accurate Knowledge</td>
</tr>
<tr>
<td>4. Press for Reasoning</td>
<td>“Why do you think that?”</td>
<td>Rigorous Thinking</td>
</tr>
<tr>
<td>5. Expand Reasoning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Take your Time</td>
<td>“Please, take your time before answering”</td>
<td>Rigorous Thinking</td>
</tr>
<tr>
<td>B. Say More</td>
<td>“That’s interesting! Can you elaborate on that?”</td>
<td></td>
</tr>
</tbody>
</table>

**Agent support for academically productive talk**

A study involving ninth-grade students as participants indicates that a Revoicing form of support (Table 1, move 2) may have a positive learning effect on novice learners, not familiar with the instructional material (Dyke et al., 2013). Additionally, a study by Adamson, Ashe, Jang, Yaron and Rosé (2013) employing university students as participants suggests that an Agree-Disagree agent intervention mode (Table 1, move 1A) can enhance students’ learning and intensify group knowledge exchange. In contrast, a similar study on university students illustrates that experienced learners may not benefit at all from a Revoicing intervention mode since they are already capable of articulating and modeling their own ideas without the need for support (Adamson et al., 2014). Following a similar rationale in higher education settings, a recent study explores the impact of a conversational agent displaying both Agree-Disagree and Add-On interventions (Table 1, moves 1A and 1B) (Tegos, Demetriadis, & Karakostas, 2015). Findings suggest that such agent interventions can improve domain-learning outcomes and increase students’ explicit reasoning during their online discussions.

The evidence obtained from the few studies conducted in this area reveals that the efficacy of the APT-based agent support can significantly vary in terms of the APT move used, the difficulty of the instructional domain, and the students’ background knowledge. This is consistent with Michaels et al.’s (2008) findings indicating that the effectiveness of APT facilitation strategies may depend on a number of factors, such as the teacher’s authority, the students’ background, and the education level. Hence, there are a number of issues remaining, far from trivial, concerning the design and use of conversational agents displaying APT-based prompts. For instance,
what is the context where each APT move can be used most effectively? Could we design an agent alternating different APT moves in accordance to the learning task or student population? What should the teacher’s role be in such a system?

Research objectives

Expanding on previous research in the utilization of agent prompts that foster accountability to the learning community (e.g., Adamson et al., 2013; Tegos et al., 2015), in the following section we present a study exploring the effectiveness of a configurable agent intervention mode promoting accountability to accurate knowledge. More specifically, the study investigates whether a series of agent interventions encouraging students to build on their prior knowledge (Table 1, move 3) can trigger explicit reasoning processes and enhance learning performance in a collaborative activity in higher education. We expect the outcome of the study to inform researchers and designers on the pedagogical value of implementing similar APT-based agent intervention modes to provide collaborative learning support and help students sustain a constructive peer dialogue.

Method

Participants and instructional domain

The study involved 72 undergraduate computer science students (18 females and 54 males). All of them were native Greek speakers and their age ranged from 19 to 23 ($M = 20.39$, $SD = 1.06$). The activity of the study was carried out in the context of a second-year course on Human-Computer Interaction (HCI) as a mandatory course assignment. All students were required to successfully complete the activity in order to pass the course. During the HCI course, students became acquainted with principles of perception and cognition required for effective interaction design (Preece, Sharp, & Rogers, 2015). Additionally, students learned useful practices for designing, prototyping, and evaluating human-centered user interfaces.

MentorChat: A configurable conversational agent system

In the study we used MentorChat, a prototype dialogue-support system (Tegos, Demetriadis, & Karakostas, 2015). As Figure 1 indicates, the interface of MentorChat, resembling an instant messaging application, asks learners to collaborate in order to accomplish one or more learning tasks in an online activity (Figure 1A). Throughout the students’ discourse, a conversational agent (Figure 1B) delivers APT-based interventions. The agent uses a text-to-speech (TTS) engine to read its interventions, which are displayed outside the main chat frame (Figure 1C).

![Figure 1. MentorChat learning environment](image-url)
MentorChat features an administration interface that allows teachers or researchers to register participants and assign them to groups, monitor students' discussions, and create collaborative learning activities. An activity may include multiple dialogue-based tasks (phases), typically asking students in each group to discuss a topic and submit a joint response to an open-ended learning question. The instructor can use the domain configuration panel (Figure 2) to model the agent domain knowledge via a concept mapping interface. A concept map can be created for each task by entering a series of simply structured statements, involving key domain concepts (Figure 2A). Each statement consists of a subject (concept X), an object (concept Y), and a verb or verbal phrase (connection). As illustrated in Figure 2B, these elements are automatically rendered and visualized in a concept map, serving as the domain knowledge representation of the agent. This authoring method facilitates the reusability of teacher-defined domain models by providing access to a library of pre-built concept maps.

Since a detailed analysis of all system components is beyond the scope of this paper, we will concisely present the functionality of the conversational agent on the basis of three core models: the peer interaction, the domain, and the intervention models (Figure 3).
dynamically updated including new concepts (nodes), as the peer dialogue unfolds. In principle, this allows the agent to direct its interventions to a specific peer (e.g., the least active). In this study, however, the agent was configured to address the group and not any individual student, simultaneously displaying its interventions to both peers.

The identification of an intervention opportunity is handled by the agent domain model. In the current version of the system, this is accomplished by comparing the nodes of the agent concept map (Figure 4A) to the nodes of the learners’ concept maps (Figure 4B). Each time a student makes a contribution, the agent looks for a domain concept that is both marked as “not yet discussed” in the agent map and included in any of the learners’ concept maps. If the above criteria are met for a concept \( \Upsilon \), the agent proposes an intervention, asking students to logically connect concept \( \Upsilon \) with an associated higher-level (parent node) domain concept \( \chi \) that is also marked as “not yet discussed” in the agent map. For example, as shown in Table 2, when a peer discusses the low-level domain concept of “mouse movement in menu” for the first time, the agent decides to intervene asking students to link the concept being discussed with a higher-level domain concept introduced by the agent (in this case, the “Accot-Zhai law”). Afterwards, the agent marks both the low- and the higher-level concept nodes as “already discussed” (Figure 4A). This mechanism ensures that no specific agent intervention is activated more than once while the agent is always aware of the domain concepts not yet discussed.

The instructor of the HCI course set up a MentorChat activity by entering a learning question asking students to collaboratively evaluate the web interface of an e-shop based on the main usability principles (Figure 1A). After
their discussion, peers were required to jointly write and submit a report emphasizing two strong and two weak aspects of the interface design, also proposing improvements in terms of efficiency and learnability. Additionally, the instructor set up the agent concept map by entering a series of statements involving key domain concepts, as depicted in Figure 2.

The participants of the study were randomly assigned to dyads and, then, to a control or a treatment group. The control group consisted of 34 students (17 dyads), whereas the treatment group consisted of 38 students (19 dyads). All students were informed that their discussions would be monitored and recorded.

The MentorChat activity took place in the university computer labs, with 40 minutes allotted for all groups. Prior to the activity students filled in a 20-minute pre-test questionnaire. Immediately after the activity, students were required to answer a post-test questionnaire as well as a student opinion questionnaire, lasting 20 and 10 minutes respectively. One week later, the treatment group students participated in a focus group session.

**Experimental design**

A randomized pre-test post-test design was adopted to assess the impact of the agent intervention mode used in this study. Two conditions were compared: (a) students who received agent interventions while collaborating in dyads to accomplish a learning task (treatment) and (b) students collaborating in dyads to accomplish the same task without any agent intervention (control).

On the basis of the peer interaction, domain, and intervention models described in the above section, the agent employed in the treatment condition delivered dynamic interventions that promoted accountability to accurate knowledge (Table 2, row 3). According to the activity guidelines, peers were expected to respond to the agent in a coordinated way (one of them) using the agent answer box (Figure 1C), which remained visible until an answer had been submitted by a student. The aim of the agent interventions was to encourage students to build on their prior knowledge, linking their current contributions to the main theoretical aspects and usability principles discussed in class.

In contrast to the treatment students, their control counterparts discussed without the presence of the conversational agent and did not receive any APT-based interventions. In both conditions, the system displayed a few static prompts that either supported group awareness (e.g., “John has logged into the activity”) or provided simple instructions on some interface features (e.g., “You can submit your answer by clicking…”).

**Data collection and analysis**

The pre-test of the study consisted of two parts. The first part included 10 multiple-choice questions requiring students to recall previously learned basic information; thus, it was relevant to Bloom’s taxonomy first level (Krathwohl, 2002). The second part included four open-ended questions asking students to demonstrate their understanding (relevant to Bloom’s taxonomy second level). Answering each of the two questionnaire parts could give students up to 10 points. Students’ answer sheets were mixed and scored by two independent raters with extensive experience in the HCI domain. Holistic rubric scales were used for the assessment of the open-ended questions. The intra-class correlation coefficient indicated a high inter-rater reliability (ICC = .98). Additionally, the internal consistency coefficient of the total pre-test scores was found to be satisfactory (Cα = .85). Since the normality and homogeneity of variance criteria were satisfied, independent samples t-tests were conducted on the pre-test scores of the control and treatment groups to compare students’ knowledge levels prior to the activity.

**Individual learning**

In order to measure individual learning, we conducted a post-test consisting of six open-ended questions. These asked students to demonstrate their comprehension of the main usability principles (second level of Bloom’s taxonomy). The sum of questionnaire items scores could give a student up to 20 points. Post-test questionnaires were assessed by the same raters as in the pre-test. Their intra-class correlation coefficient was found to be high (ICC = .98). An analysis of covariate (ANCOVA) was performed on post-test scores to explore the difference in individual learning outcomes in the two conditions.
**Group learning**

In order to measure group learning, the raters assessed the quality of the dyads’ answers submitted at the end of the activity to the learning question (Figure 1A). The group answers were scored using a rubric rating scale while a high inter-rater reliability was obtained (ICC = .97). An independent samples t-test was conducted to compare the group learning outcomes under the control and treatment conditions.

**Explicitness**

In order to measure the explicitness of students’ reasoning, a discourse analysis was performed by the authors. The analysis was based on an adjusted version of the IBIS discussion model widely considered to be an effective framework for analyzing the conversational interactions occurring in remote collaboration sessions (Liu & Tsai, 2008). Drawing on the main categories of the IBIS model (issue, position, and argument), our scheme introduced two additional categories, named explicit position and explicit argument, both focusing on the display of students’ reasoning on domain concepts. Table 3 presents the final scheme used for coding students’ contributions, each of which could consist of multiple chat posts. Mann Whitney U-tests evaluated the difference in explicitness between the conditions by comparing the explicit position and explicit argument frequencies of the control and treatment dyads.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off-task</td>
<td>Utterances playing a purely social function or not relating to the task (e.g., “Greetings”)</td>
</tr>
<tr>
<td>Repetition</td>
<td>Reiterations of prior contributions</td>
</tr>
<tr>
<td>Management</td>
<td>Management-oriented statements used for task coordination (e.g., “Let’s submit our answer”)</td>
</tr>
<tr>
<td>Common Understanding</td>
<td>Short (typically one- or two-word) utterances establishing common understanding on the subject (e.g., “OK,” “I see”)</td>
</tr>
<tr>
<td>Issue</td>
<td>What needs to be done or resolved to proceed with the overall task (e.g., “What other laws are relevant?”)</td>
</tr>
<tr>
<td>Position</td>
<td>Opinions usually related to the resolution of the issue raised (e.g., “Fitts’ law applies here”)</td>
</tr>
<tr>
<td>Argument</td>
<td>Opinions supporting or objecting to a position (e.g., “You are absolutely right”)</td>
</tr>
<tr>
<td>Explicit Position</td>
<td>Positions explicitly displaying reasoning on domain concepts (e.g., “According to Hick-Hyman, the reaction time increases logarithmically as the number of options increases”)</td>
</tr>
<tr>
<td>Explicit Argument</td>
<td>Much as explicit positions, arguments displaying explicit reasoning on domain concepts (e.g., “I disagree, Hick’s law cannot be used for randomly ordered lists requiring linear time”)</td>
</tr>
</tbody>
</table>

**Explicit response ratio**

In order to gain a deeper insight into how the intervention mode affected the students’ explicitness, an “explicit response ratio” (ERR) was calculated by dividing the number of agent-induced explicit contributions (explicit positions and explicit arguments) by the number of agent interventions. An explicit position or explicit argument was labelled as “agent-induced” only if it was a direct response to the agent or a subsequent comment relating to the agent intervention (Table 2, rows 5-7). Thus, the ERR value of a dyad could range from zero, if the agent interventions had no effect on the generation of explicit contributions, to more than one, when multiple explicit contributions were triggered by each agent intervention.

Following an exploratory content analysis drawing inferences on peers’ behavior after an agent intervention, a mediation analysis was implemented. The analysis investigated whether the generation of explicit positions and explicit arguments in students’ dialogue could account for a potential association between the conditions and the learning outcomes. In general, a variable is regarded as a mediator if it carries the influence of a given independent variable to a given dependent variable (Rucker, Preacher, Tormala, & Petty, 2011). In our study, the intervention mode (activated in the treatment condition and deactivated in the control condition) was the independent variable, while individual learning (as measured by the post-test) was the dependent variable. Hence, our analysis tested the mediating effect of explicitness (as measured by students’ explicit position and explicit argument frequencies) on the relationship between the agent intervention mode and the learning outcomes. Using the PROCESS SPSS macro, the test was performed by applying bootstrapping with bias-
corrected confidence estimates (Hayes & Preacher, 2014). The 95% confidence interval of the indirect effects was obtained with 5000 bootstrap resamples.

**Students’ opinions**

The student opinion questionnaire included Likert-scale questions ranging from 1 (disagree) to 5 (agree). The first questionnaire part elicited students’ opinions regarding their learning experience in MentorChat ($C_α = .86$). A Mann Whitney U-test was conducted for each questionnaire variable to assess the difference between the scores reported in the two conditions. The second questionnaire part consisted of four pairs of positively and negatively keyed questionnaire items, which were only available for the treatment students and measured their perceptions of the agent interventions. After reverse scoring all negatively keyed items, we calculated a mean score for each question pair ($C_α > .70$ for all pairs).

The focus group session followed a semi-structured protocol and explored the opinions of the treatment students on the display of the interventions and their interaction with the conversational agent. Students’ responses were transcribed verbatim and analyzed in search of common themes following an open coding process.

**Results**

In respect to the closed-type pre-test questions, the independent-samples t-test yielded no statistically significant differences as the treatment group mean ($n = 38$, $M = 5.89$, $SD = 1.59$) was comparable to that of the control group ($n = 34$, $M = 5.97$, $SD = 1.96$), $t(70) = .181$, $p = .857$, $d = .043$. Likewise, the treatment students ($n = 38$, $M = 4.49$, $SD = 2.28$) performed similarly to control students ($n = 34$, $M = 4.35$, $SD = 2.16$) in the open-ended questions, $t(70) = .260$, $p = .795$, $d = .062$.

**Individual learning**

The results of the ANCOVA conducted on post-test scores, with total pre-test scores as a covariate, revealed a significant difference in favor of the treatment group. The analysis showed that the students in the treatment condition ($n = 38$, $M = 12.12$, $SD = 4.41$) outperformed their control counterparts ($n = 34$, $M = 9.45$, $SD = 3.76$) on individual learning, $F(1,69) = 9.162$, $p = .003$, $η_p^2 = .117$.

**Group learning**

The independent-samples t-test evaluating group learning indicated that the dyads in the treatment condition achieved considerably better scores ($n = 19$, $M = 14.03$, $SD = 3.46$) than the dyads in the control condition ($n = 17$, $M = 10.56$, $SD = 3.93$), $t(34) = 2.814$, $p = .008$, $d = .965$.

**Explicitness**

The discourse analysis identified 2646 students’ contributions. From those 1543 came from the treatment group while 1103 came from the control group. The overall results of the coding process are summarized in Table 4.

<table>
<thead>
<tr>
<th>Category</th>
<th>Control (n = 17 dyads)</th>
<th>Treatment (n = 19 dyads)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Frequency (%)</td>
<td>Group average</td>
</tr>
<tr>
<td>Off-task</td>
<td>85</td>
<td>7.71</td>
</tr>
<tr>
<td>Repetition</td>
<td>28</td>
<td>2.54</td>
</tr>
<tr>
<td>Management</td>
<td>187</td>
<td>16.95</td>
</tr>
<tr>
<td>Common Understanding</td>
<td>160</td>
<td>14.51</td>
</tr>
<tr>
<td>Issue</td>
<td>93</td>
<td>8.43</td>
</tr>
<tr>
<td>Position</td>
<td>202</td>
<td>18.31</td>
</tr>
<tr>
<td>Argument</td>
<td>201</td>
<td>18.22</td>
</tr>
</tbody>
</table>
A Mann Whitney U-test revealed that the treatment groups \((n = 19, M = 13.42, SD = 4.55)\) generated substantially more explicit positions than the control groups \((n = 17, M = 9.22, SD = 5.22)\), \(U = 243.5, p = .008, r = .433\). The treatment groups produced also considerably more explicit arguments \((n = 19, M = 7.48, SD = 3.32)\) compared to the control groups \((n = 17, M = 4.32, SD = 2.26)\), \(U = 258, p = .002, r = .510\).

### Explicit response ratio

The average explicit response ratio (ERR) is presented in the last row of Table 5, which depicts the results of the second-level analysis focusing on the agent effect on the students’ explicitness.

<table>
<thead>
<tr>
<th>Category</th>
<th>Treatment ((n = 19) dyads)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent Interventions</td>
<td>(M = 2.74, SD = 0.73)</td>
</tr>
<tr>
<td>Agent-Induced Explicit Positions</td>
<td>(M = 4.21, SD = 1.96)</td>
</tr>
<tr>
<td>Agent-Induced Explicit Arguments</td>
<td>(M = 3.05, SD = 2.22)</td>
</tr>
<tr>
<td>Explicit Response Ratio (ERR)</td>
<td>(M = 2.69, SD = 1.19)</td>
</tr>
</tbody>
</table>

A preliminary content analysis revealed that in 16 out of the 19 dyads \((F = 84.21\%)\) the majority of the interventions were addressed by the most active group member, the one with the highest rate of explicit contributions. Although students’ responded to the agent using clear statements, a lack of peer coordination was identified following the onscreen display of some interventions \((F = 15.38\%)\), when one student responded to the agent before communicating with their partner. Additionally, it was noted that only 2 agent interventions \((F = 3.85\%)\) were ignored by the students. Both interventions were displayed towards the end of the activity, when the students had already begun composing their final group answer.

A regression analysis investigated whether the generation of explicit positions and explicit arguments mediated the effect of the agent intervention mode (treatment condition) on the post-test scores. The results showed that the agent intervention mode was a significant predictor of both students’ explicitness \((B = 3.659, t(70) = 3.197, p = .002)\) and individual learning \((B = 2.668, t(70) = 2.744, p = .008)\), while explicitness was a significant predictor of individual learning \((B = 0.291, t(70) = 3.028, p = .004)\). The mediation analysis also resulted in a statistically significant indirect coefficient \((B = 1.07; CI = .280 to 2.420)\), thus supporting the mediating role of explicitness (Figure 5). In fact, a full mediation was suggested since the agent intervention mode was no longer a significant predictor of individual learning outcomes after controlling for the mediator \((B = 1.603, t(70) = 1.628, p = .108)\), i.e., the frequency of explicit positions and arguments (Rucker et al., 2011).

![Figure 5. Mediation diagram (Note. *p < .01)](image)

### Students’ opinions

Table 6 depicts the results of two questionnaire items assessing the subjectively perceived benefits of the collaborative activity. Other questions on the interface usability revealed no major issues regarding the ease of
use \( n = 72, M = 4.07, SD = .70 \) or the performance \( n = 72, M = 4.36, SD = .66 \) of MentorChat in either condition.

<table>
<thead>
<tr>
<th>Questionnaire item</th>
<th>Control ( (n = 34) )</th>
<th>Treatment ( (n = 38) )</th>
<th>Mann-Whitney U-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The collaborative activity improved my domain knowledge</td>
<td>( M = 3.44, SD = 1.16 )</td>
<td>( M = 3.97, SD = 0.68 )</td>
<td>( U = 825, p = .033^*, r = .251 )</td>
</tr>
<tr>
<td>2. The collaborative activity was beneficial for me (regardless of any improvement in my domain knowledge)</td>
<td>( M = 4.00, SD = 1.06 )</td>
<td>( M = 4.26, SD = 0.65 )</td>
<td>( U = 711, p = .429, r = .093 )</td>
</tr>
</tbody>
</table>

Note. \(^*p < .05\).

In respect to the students’ opinions about the conversational agent interventions, Table 7 presents the total scale scores on the four questionnaire measures.

<table>
<thead>
<tr>
<th>Questionnaire item</th>
<th>Treatment ( (n = 38) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The agent questions displayed during the discussion were simple and understandable</td>
<td>( M = 4.53, SD = 0.60 )</td>
</tr>
<tr>
<td>2. The agent questions helped me recall or retrieve useful domain information for the evaluation of the e-shop interface</td>
<td>( M = 4.03, SD = 0.88 )</td>
</tr>
<tr>
<td>3. The agent questions did not disrupt my discussion with my partner</td>
<td>( M = 3.95, SD = 0.96 )</td>
</tr>
<tr>
<td>4. The timing and the content of the agent questions were consistent with the on-going discussion</td>
<td>( M = 4.08, SD = 0.78 )</td>
</tr>
</tbody>
</table>

The analysis of the focus group data resulted in the identification of the following themes: (a) the agent interventions appeared to have helped many students to focus their discussion on important usability principles \( (F = 79\%) \), (b) some students appreciated the fact that the agent did not intervene very frequently and allowed time for their response \( (F = 39\%) \), and (c) the background color of the agent answer box was perceived as disruptive by a few participants \( (F = 16\%) \).

**Discussion**

The assessment of students’ post-test answers revealed that the students of the treatment group achieved higher learning outcomes as compared to the students of the control group. Even though the knowledge levels of the students did not differ significantly prior to the activity, the treatment students outperformed their control counterparts in understanding and illustrating conceptual domain knowledge following the activity. Furthermore, when asked whether the collaborative activity improved their domain knowledge, the treatment students were more positive than the control ones (Table 6, item 1). The agent effect on individual learning appears to accord with Adamson et al.’s (2013) results, which showed that the introduction of an agent displaying Agree/Disagree interventions (Table 1, move 1A) could improve learning and intensify the exchange of ideas in higher education.

As for the group learning, it was shown that the final answers of the treatment dyads were superior to those of the control dyads in terms of correctness and comprehensiveness. Indeed, the answers submitted in the treatment condition appeared to be more conceptually complete, including several appropriate references to theoretical principles of usability. The finding is consistent with our previous findings demonstrating that APT-based agent interventions can effectively trigger group discussions to elaborate on more domain concepts, thus leading to more comprehensive group answers (Tegos et al., 2015).

The above evidence suggests that agent interventions had a positive effect on both individual and group learning. This is also supported by students’ views expressed in the student opinion questionnaire and the focus group session. Many students commented favorably on the conversational agent interventions, which they believed to have assisted them in recalling and focusing their discussions on central domain principles (e.g., Table 7, item 2). Despite the relatively simple intervention mechanism, we argue that agent interventions effectively scaffolded students’ discussions and kept students focused by asking them to support their contributions using fundamental conceptual knowledge, known to be critical to understanding and resolving learning tasks (Streveler, Litzinger, Miller, & Steif, 2008).
The discourse analysis yielded a series of noteworthy results. First, the agent intervention mode was shown to considerably impact on the explicitness of students’ reasoning. In particular, the frequency of explicit positions and explicit arguments was found to be significantly higher in the treatment group than the control group (Table 4, rows 8 and 9). The explicit reasoning enhancement seems to be owed to the agent interventions, since the number of explicit contributions identified as agent-induced (Table 5, rows 2 and 3) can justify the difference in the generation of explicit positions and explicit arguments between the conditions (Table 4, rows 8 and 9).

The explicit-response ratio (ERR), measuring the effect of the intervention mode on explicitness, was found to be higher in comparison to our previous studies (e.g., Tegos et al., 2014; Tegos et al., 2015). Although we cannot directly compare the ERR values of different studies due to their dissimilar contexts, the average ERR value reported in this study (Table 5, row 4) suggests that each agent intervention stimulated more than two students’ explicit contributions. We argue that agent interventions promoted students’ sound reasoning by urging for clear and compelling statements supported by concrete evidence. Students were encouraged to seek out appropriate evidence in order to strengthen their claims with specific domain principles and explicate their reasoning, thus making it available for others (or themselves) to assess, question or challenge (Michaels et al., 2010). This lies in the core of transactivity theory, which highlights the value of explicit reasoning displays, so that the reasoning can be communicated to the partner(s) and then built upon by the latter (Stahl & Rosé, 2011). This kind of explicitness is considered to be a matter of importance in written discussions, where learners may not explicate their thinking to the group, therefore obstructing the negotiation of common ground (Weinberger, Stegmann, & Fischer, 2007).

As illustrated by the path diagram in Figure 5, the influence of the agent interventions on students’ explicitness played a significant mediating role. A mediation analysis showed that the impact of the agent intervention mode on individual learning was mediated by the stimulation of explicit reasoning during students’ discussions. This partially confirms a previous finding showing that the effectiveness of an APT-based intervention mode depends on its ability to engage students in transactive conversation interactions such as explicit argumentation, where students communicate their reasoning and build on each other’s ideas (Tegos et al., 2015). Moreover, there is some converging evidence suggesting that the frequency of students’ contributions containing a form of explicit reasoning can act as a robust predictor of students’ learning in dialogue-based collaborative learning activities.

As opposed to the directed intervention mode used in our previous work (Tegos et al., 2015), which addressed a specific student of the group, the agent of this study allowed the peers to decide who would respond to its interaction by themselves. While the activity guidelines clearly stated that peers should discuss and coordinate their answers to the agent, we observed that the most active group members, those having the highest explicit position and explicit argument frequencies in their dyads, often took the initiative to respond to the agent before communicating with their partners. Based on prior findings on the effectiveness of differently presented intervention techniques (Tegos et al., 2014), we believe this miscommunication between peers to have resulted from the reduced situational constraints, i.e., not requiring students to follow a specific protocol in their interaction with the agent. Although further investigation is needed to draw definite conclusions about how group conversations were affected, we consider that the undirected nature of the agent interventions shifted the overall balance of students’ discussions in the treatment condition by allowing the most active peers to regulate the student-agent interaction by themselves.

The relatively low number of interventions displayed in the treatment condition (Table 5, row 1) was commented favorably in the focus group discussion, as many students appreciated the non-frequent intervention mechanism of the agent, which did not appear to have a major interruption effect based on our observations and students’ opinions (Table 7, item 3). Additionally, our content analysis revealed that the agent effectively drew students’ attention since only two interventions were ignored by the students, who generally perceived the agent interventions as quite comprehensible and consistent with the ongoing discussion topic (Table 7, items 1 and 4). Yet, in interpreting the above findings one should consider the study limitations as well. First, the participants were aware of their discussions being monitored and recorded, which could have altered the typical conversational behavior of students, as they may have paid more attention to the agent interventions than they would have in a non-controlled environment. Second, due to our limited sample size, study findings need to be confirmed by a larger sample size and other student populations of different backgrounds or ages. Lastly, it is clear that the agent does not possess intelligence capable of engaging into a full-fledged discussion with students. Still, this is in line with our objective of building easily deployed agents that can accomplish substantial learning benefits with the minimum required level of intelligence.
Conclusion

The study provides promising evidence on the utilization of a teacher-configurable agent intervention mode that encourages peers to build on their prior knowledge, drawing on the academically productive talk (APT) discourse framework. Study findings indicate that agent interventions aiming to link students’ contributions to previously acquired knowledge can improve both individual and group learning when implemented in the context of a collaborative learning activity in higher education. The agent interventions have also a positive impact on students’ conversational behavior by amplifying explicit reasoning. In fact, the level of explicit reasoning appears to mediate the relationship between the agent interventions and individual learning. Overall, we believe the results of this study to contribute to the understanding of how the facilitation of collaborating groups can be effectively automated through agile agent interventions. In the future, we aspire to utilize our configurable conversational agent system to explore the impact of other APT facilitation strategies in various educational domains and levels.

References


Journalogue: Voicing Student Challenges in Writing through a Classroom Blog

Suneeta Thomas
Department of English, Missouri State University, USA // suneetathomas@missouristate.edu, sthomasacademic@gmail.com

(Submitted October 10, 2015; Revised February 15, 2016; Accepted 24/04/2016)

ABSTRACT
This study qualitatively analyzes the challenges presented by international undergraduate students in a freshman composition course at a large Midwestern university in the US. 15 students were divided into five groups, of three members each, with varying proficiency levels in writing. They were asked to submit reflections as journal/blog posts, on two assignments (each, a different genre). Each group was then asked to comment on their group members’ reflections, while the teacher remained a silent observer, to ensure a stress-free environment. With a total of 30 blog posts, and 60 comments, the data was qualitatively coded to display major themes. It was found that the journal/blog posts described student challenges in writing, but also shared optimism for learner growth. Collaborative reflection and learning was accomplished through students’ comments, which irrespective of language proficiency, showed identification with each other, and provided encouragement, and advice on how to deal with these challenges. Additionally, this online mode of expression allowed complete teacher access to understand linguistic challenges of students so as to make necessary alterations in the classroom. The paper concludes by presenting suggestions that can be pedagogically used to address the problems faced by students, and by delineating avenues for further research.

Keywords
Classroom blogging, ESL writing, Identity, Reflective learning, Freshman composition

Introduction

With rising access to technologies in the classroom, discussions on blogging and its learning effects have evolved in second language writing studies in recent years. Blogging, as a pedagogical tool has shown to enhance learner engagement, foster knowledge, and increase socio-cultural interaction in the classroom. Specifically, as previous scholars have observed, blogs are significant due to characteristics such as, encouraging communicative and reflective practices (Absalom & De Saint Léger, 2011; Bhattacharya & Chauhan, 2010; Sidhu, Kaur, & Fook, 2010), motivating students to write, improving writer performance (Arslan & Ṣahin-Kızıl, 2010; Lee, 2010; Lin, Li, Hung, & Huang, 2014), its access in real-time, and creation of personal ownership in a public space (Absalom & De Saint Léger, 2011; Chen, 2015; Lin et al., 2014; Sun, 2009; Sun & Chang, 2012). Also referred to as the Blog Assisted Language Learning (BALL) approach (Lin, Groom, & Lin, 2013), Chen (2015) observes that the literature has shown to have three strands of pedagogical foci, namely the development of: (1) linguistic skills, (2) metacognitive skills, and (3) intercultural competence. For instance, with regard to linguistic skills, Fellner and Apple (2006) discovered an increase in writing fluency in their students, by the end of their study, than when they had begun. Similar studies include Bloch (2007), and Gebhard, Shin and Seger (2011) who examined the development of rhetorical strategies, and syntactic complexity in their subjects, respectively.

In terms of metacognitive skills, scholars have discovered that blogging urges students to think about their thinking (to reflect), and to write about their writing. That is, students are asked to actively monitor and evaluate the production of their writing, outlining specific techniques and strategies used during the process. A good example is Sun and Chang’s (2012) study that investigates EFL graduate students’ blogging efforts in reflecting on their writing processes and identities. The authors show how collaborative dialogue between students can reconstruct writer knowledge, and identity. Other studies along this line of research include Absalom and De Saint Léger (2011), Mompean (2010), and Bhattacharya and Chauhan (2010), among others who investigate the role of reflection, motivation, and autonomy in ESL/EFL students’ blogging process. The third strand of research, categorized as the development of intercultural competence has been documented in contexts of study-abroad programs, and in scenarios where students interacted with native speakers of the target language, or learners from different linguistic, and cultural backgrounds. For instance, Ducate and Lomicka (2005) in a German FL learning context, observed high motivation in students wanting to relate to the target culture. This was because they had required their students to research, compare, and write about German culture on their blogs, thus encouraging critical analytical skills, as opposed to being passive learners in the classroom. Similarly, Melo-Pfeifer (2015) in her research on Portuguese foreign language learning, investigated the use of a
pedagogical blog to facilitate plurilingual and intercultural competences in her students. Her study revealed that her students by the end of the project had improved their vocabulary, and cultural knowledge of the target language. Other studies such as Elola and Oskoz (2008), Comas-Quinn, Mardomingo, and Valentine (2009), Yang (2011), and Garcia-Sanchez and Rojas-Lizana (2012), also show learner engagement with the target culture, and the development of target culture awareness. However, as much as the pedagogical benefits of blogging are visible from the above studies, a new emerging fourth strand reveals some drawbacks too. Scholars such as Chen (2015), Lin (2014), Lin et al. (2014), Lin et al. (2013), and Lin et al. (2011), for instance, have come to grips with the inevitability of blogging activities ceasing once the writing course ended. Nevertheless, even with this setback, the benefits of blogging on writing and learning, far out-weigh the cons.

While a majority of L2 blogging studies have been conducted in EFL settings, minimal research has been conducted in North American ESL freshman composition classrooms (cf. Bloch, 2007). This gap is significant because international students who attend undergraduate studies in the United States (US) are in many ways similar to student populations in ESL/EFL settings. A majority of these students are fresh out of high schools from their ESL/EFL backgrounds. Yet, unlike previous studies, they experience different challenges in linguistic skills, particularly in writing since it is in a US composition classroom context. This study addresses one aspect of their writing challenges by investigating the blog use of international students in an American composition classroom. Adopting an expressivist approach with elements of socio-cognitivism, this study merges the concepts of journaling, blogging, and dialogue to theoretically inform this study. The title “Journalogue,” as a result, emerged from the merging of these three concepts.

Theoretical framework

Expressivism or the expressivist approach, according to Walter-Echols (2008), began as a criticism towards “impersonal and product-oriented” tasks in L1 composition in the 1970s (p. 121). Instead of impersonal tasks, the expressivist approach sought to express the personality of the author through his/her voice. It sought to be a creative outlet, engaging in self-expression, and self-discovery of the writer. Journaling, as a consequence, emerged as a method to absolve this need. Journaling prioritizes the act of reflection, and has been referred to as, reflective writing, or maintaining reflective journals, learning logs, learning journals, research logs, diary entries, etc. (Moon, 2006). Thorpe (2004) in defining reflection, quotes Boyd and Fales’ (1983) who state that it is “the process of internally examining and exploring an issue of concern, triggered by an experience, which creates and clarifies meaning in terms of self and which results in a changed conceptual perspective” (p. 100, as cited in Thorpe 2004, p. 328). When it comes to ESL writers, Sidhu et al. (2010) for example, study how students monitor their learning through journal writing. Veerappan, Suan, and Sulaiman (2011) further look at the effect of teacher scaffolding during the journaling process. More importantly, Sidhu et al. (2010) advocate “Writing reflectively about their language learning experience is at the heart of ESL students’ experience and as such evaluation of students’ experience defines what students regard as important and how they come to see themselves developing academically as language learners” (p. 128). While reflection is identified as a dialogue with oneself, a blog expands this visibility of self-dialogue to a broader public space, encouraging public participation, and consumption through its comment feature. The intersection of reflective journaling, blogging (expressivism), and consequent dialogue and interaction (socio-cognitivism), aka creation of journalogues, forms the crux of this study. In this context, the teacher is a passive observer, while students participate in interaction and dialogue through the medium of a blog. In their participation, students are not solving a hands-on problem, or conducting a collaborative project per se. Rather, adopting the rudimentary elements of the socio-cognitivist approach (Murray & Hourigan, 2008), this study observes student reflection of challenges in writing, while peers engage in a consequent cognitive discussion.

Methodology

Larger context of the study

The United States (US) has been seeing a tremendous surge of international students from ESL/EFL backgrounds in university settings (Jordan, 2015). Although, to characterize this international population in its entirety would be incautious and beyond the scope of this paper, it is safe to say that a majority of these students, especially those from EFL countries require support in getting assimilated into the American classroom. An important aspect of such support is language learning, and assimilation. When international students come to pursue undergraduate studies in the US, some sources of support available to them are ESL conversation groups, or enrollment in intensive English language programs. However, they are expected to take a freshman
composition course (FYC) which is a mandatory writing course requirement for degree completion. Depending on the university, they are either placed in the mainstream composition course with domestic, native speaker students, or some universities allow the option for choosing an equal, but an ESL equivalent FYC. The context of the current study, involves the latter: International undergraduate students registered for an international section of FYC in a large Midwestern university.

The study: Curriculum design

The present study, seeks to analyze students’ reflective journal/blog posts that describe their challenges in writing specific assignments. One section of an international freshman composition classroom, comprising a total of 15 students, taught by the researcher were chosen as participants for this study. The curriculum of the course requires students to participate in a sequenced writing project. That is, they write a sum of four papers on the same subject over the course of the semester. Completing a series of assignments for a sequenced writing project represents the belief that students improve their writing skillset when each assignment builds on the experience and knowledge gained from completing previous assignments (Leki, 1992). In the sequenced writing project, the students may write on any topic they wish with the approval of the instructor. However, it must be a topic that will allow them to do all four assignments (Personal Narrative, Interview Report, Literature Review, and Argumentative Essay) of the project. Additionally, each assignment undergoes three revisions, and only the final, third draft is graded. This method allows the students to revise their papers twice (or even more should they choose to) before they submit their final. This is how the course is designed.

Participants

The participants were students enrolled in one section of an ESL writing course in a large Midwestern university during the Spring 2014 semester. The class had mostly freshmen, with three sophomores, and one junior student. Since it was the spring semester, the students had had the chance of experiencing the university for at least one semester (in the preceding Fall). Thus, all were familiar with the university environment which predisposed them to a less stressful semester, in general. Additionally, the researcher was fortunate to have a diverse group of students in the class with six Chinese students, four Indian students, three Malaysian students, one Honduran, and one Korean student. Table 1 shows the demographic information of these participants.

<table>
<thead>
<tr>
<th>Students</th>
<th>Languages spoken</th>
<th>Duration of English studied (average)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Chinese</td>
<td>Mandarin, Wu Chinese</td>
<td>7.3</td>
</tr>
<tr>
<td>4 Indian</td>
<td>Hindi, Marathi, Sindhi, French</td>
<td>11.5</td>
</tr>
<tr>
<td>3 Malaysian</td>
<td>Malay, Mandarin</td>
<td>15</td>
</tr>
<tr>
<td>1 Honduran</td>
<td>Spanish, French</td>
<td>15+</td>
</tr>
<tr>
<td>1 Korean</td>
<td>Korean</td>
<td>10</td>
</tr>
</tbody>
</table>

Procedure and data analysis

For the purposes of this study, at the end of each assignment cycle once the students submitted the final draft, they were asked to write a (300+ words) reflection on their writing experience of that assignment (See Appendix A). They were provided a separate day, typically the day after their final submission, to carry out the journal assignment in class. In this way, the writing experience was still fresh in their minds, and they were less stressed since the final draft had already been submitted. The journal posts were done online and submitted through the university Blackboard learning system. The posts were designed such that it was visible only to the students registered in the class. Additionally, each student was asked to comment (100+ words) on two fellow-group-member’s journal posts of the same assignment (See Appendix B). In order to accurately assign group membership, students were assigned one of the three writing proficiency levels—high (H), medium (M), or low (L) based on their performance on a written diagnostic test, and subsequent drafts of the first assignment. For analysis, it is best understood if the students are visualized on a cline of proficiency as seen in Table 2. Five students were placed under each proficiency level and were assigned code names. For instance, in the high proficiency level, the top five students were assigned H1, through H5 where the digit 1 represented the highest proficient student, while the digit 5 represented the least proficient student. In order to ensure that each student was getting feedback from students of other proficiency levels, they were subsequently placed in groups of three,
each comprising of a high, medium and a low level student. They were grouped as follows: H1, M1, L1 in group 1; H2, M2, L2 in group 2; and so on.

<table>
<thead>
<tr>
<th>High</th>
<th>Medium</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1, H2, H3, H4, H5</td>
<td>M1, M2, M3, M4, M5</td>
<td>L1, L2, L3, L4, L5</td>
</tr>
</tbody>
</table>

Table 2. Cline of proficiency

For the purposes of this paper, journal posts on two assignments—Personal narrative and Literature Review were assessed. This led to a total of 15 x 2 journal posts (30 journal posts) and 2 comments per post (60 total comments). The data was then qualitatively analyzed through coding and memoing and quantified to display major themes in the reflections and comments.

In developing the research design, three main research questions guided this study:
(1) What kind of writing challenges do students share in their journal posts?
(2) How do fellow group members respond to their group members’ posts?
(3) What is the nature of interaction between different proficiency level students through the commenting feature?

Results

Journal codes

Table 3 briefly summarizes all the different journal codes that were found after coding. Here, we will look at a few of the prominent and distinguishable ones. Out of the challenges (C-) that affected all students, linguistic challenges seem to have topped the list, with 26 tokens (occurrences) mentioned in the Personal Narrative (PN) journal posts, followed by 18 tokens mentioned in the Literature Review (LR) journal posts. These tokens represented challenges in relation to grammar, vocabulary, semantics, and mechanics. The next highest occurring code was the process/strategy code with 21 tokens. This code represented the strategies students employed in order to cope with or tackle the different challenges (including linguistic challenges) they faced in the assignment. Accomplishments ranked third, with 11 tokens featuring in the PN journal and only two tokens in the LR journal. In line with Accomplishments was also the code of Sources that referred to the difficulty in choosing the right sources for the LR journal. Since this code featured only in the literature review assignment and occurred 13 times, it goes on to show the high percentage it had in terms of challenging the students. Some students even mentioned the challenge of either going under or over the word limit, with five tokens featuring in the PN journal and only three tokens in the LR journal. Another important challenge that the students mentioned in their posts was about citations. Many students shared the difficulties of citing properly while a handful were happy that they were able to easily follow the format and do them. Although there were other challenges mentioned in the journal posts, these reflections did not go without any positivity about the writing experience. There were five tokens on the code Writing is fun, and five that were colored with hope and positivity. Both these codes came from the PN journal showing the enthusiasm of the students in their first assignment of the course. The final code that I would like to discuss is the lack of confidence code. Although there were only two tokens of this code, it showcased the true emotions and fears of the student. All these codes are discussed in detail in the analysis section.

Comment codes

The comments that were entered after these posts showed four major themes/codes: Empathy, encouragement, compliments, advice/suggestions. The Empathy code was recorded whenever students showed instances of relating to the sentiment of the journal-author, or shared the same experiences as the journal-author. Encouragement code referred to instances where students provided encouragement to strive better or to continue with the journal-author’s present endeavors. Compliments code referred to instances where the student expressed awe and complimented the journal-author for his/her accomplishments or perseverance. Finally, the advice/suggestions code referred to students offering suggestions for help, or advice in response to some of the challenges mentioned in the posts. Although on a superficial level, it may appear that these codes overlap, strict measures were taken to ensure that the codes were represented correctly in the data. Additionally, since there was tremendous repetition of these codes in all of the 60 comments, they were not quantified. Rather the form of interaction between the different proficiency level students was analyzed. This will be discussed in the next section.
Table 3. Journal codes

<table>
<thead>
<tr>
<th>Codes</th>
<th>Description</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-Linguistic</td>
<td>Grammar, mechanics, semantics</td>
<td>26 + 18 (44)</td>
</tr>
<tr>
<td>Process/strategy</td>
<td>Coping with challenges</td>
<td>21</td>
</tr>
<tr>
<td>Accomplishments</td>
<td>Writing, language, expression, research</td>
<td>11 + 2 (13)</td>
</tr>
<tr>
<td>Sources</td>
<td>Searching for good sources</td>
<td>13</td>
</tr>
<tr>
<td>C- Word limit</td>
<td>Less/excess</td>
<td>5 + 3 (8)</td>
</tr>
<tr>
<td>Citation</td>
<td>Struggles/Ease</td>
<td>7</td>
</tr>
<tr>
<td>C-Topic choice</td>
<td>Difficulty choosing topics</td>
<td>6</td>
</tr>
<tr>
<td>Writing is fun</td>
<td>Enjoyment</td>
<td>5</td>
</tr>
<tr>
<td>Positive/hope</td>
<td>Hope for better writing</td>
<td>5</td>
</tr>
<tr>
<td>Lack of confidence</td>
<td>Lack of confidence</td>
<td>3</td>
</tr>
<tr>
<td>L1/L2 translation</td>
<td>Translation issues</td>
<td>3</td>
</tr>
<tr>
<td>Writing is difficult</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>American teaching</td>
<td>Preference for ideas than grammar</td>
<td>2</td>
</tr>
<tr>
<td>C- Research Qs</td>
<td>Formulating Research Qs</td>
<td>2</td>
</tr>
<tr>
<td>Personal exp.</td>
<td>Description of exp.</td>
<td>2</td>
</tr>
</tbody>
</table>

Qualitative analysis

Journal codes

Following are some excerpts and examples of the above journal codes:

*C-Linguistic*

“...I actually do spend a lofty amount of time deciding how to phrase my sentences. I would try to play around with words, crafting bombastic and colorful sentences. ‘How much of these could I use before crossing the line of sounding like nonsensical ramblings?’ I was also tentative to how my sentences would read out. I had rephrased the same line, over and over again, out of frustration, trying to rectify how awkward it would sound.” - H1

The above H1 student, the highest performing student in the class has impressive literary writing skills. In the above journal excerpt, he clarifies that he spends a lot of time trying to out-do himself in his writing in order to do well in this class. Despite his advanced writing skills, he too is faced with linguistic challenges such as making appropriate word choices in his writing.

“When S (instructor) revised my essay I understood all my errors and those were mostly the run-on and repeated sentences and some of things in essay were explained properly. So, I have to work a lot to improve my essay, I found it hard because there were errors in almost all my essay and I was difficult for me to correct my essay. But somehow I did it. But after the revision of my second draft, I found more errors and they were the same, but this time it was run-on sentences and repeated words.” - M3

The above excerpt by a medium level student shows his linguistic challenge of writing run-on sentences. His writing style is starkly different from H1, as is visible through the run-on sentence he uses within this excerpt.

“When I write my essay, I meet many problems. Such as my grammar is not very good, I cannot remember many words and I do not know how to write so much words. This is my first year come to United States, so my English is not very good. I look up the words in the dictionary, ask my roommate about my grammar. I try my best to write the essay better.” - L5

This excerpt highlights the linguistic struggles of the lowest level student in the class. Once again showing a strong difference between the above two excerpts, this excerpt discloses a heightened degree of linguistic challenges than a medium or a high level student. In a similar fashion, students also mentioned challenges in relation to grammar, mechanics and semantics that constituted the Linguistic challenges code.
Accomplishments

“Before doing this assignment, I only know about this topic on the surface, but I am so glad that I chose this topic because I found a lot of new information regarding this crucial issue that is affecting the society. I also gained a whole new level of understanding about the media and its impact on the world!” -H2

This is by the second highest performing student who discusses the many accomplishments he has achieved while writing his personal narrative assignment. Similar instances were seen in the medium and low level students, as well.

“I learned how to analyze the requirements of the essay if it is offered and that can help me write an essay with an appropriate topic related to the requirements. Also, if there is no specific requirement, I know how to select a topic that I am interested in and will can inspire me. Another progress for my writing skills is I learned some new skills about research. This class taught me how to expand my topic to find the problem I need to research.” -M4

The positivity and enthusiasm of the student comes through, through this excerpt on the accomplishments he has had after one month of writing.

“Since the beginning of the new semester, I have spent one month on the writing class. I also learned a lot during this time… Although I spent much time on this assignment, I still feel so happy cuz I gain a lot. I know how to reach the agreement of the teacher, the format of the article and so on.” -L4

Although by a low level student, it was very encouraging to see that he too had felt a sense of accomplishment in his writing. However, we can notice the student mentions the idea of “agreeing with the teacher” as a necessary quality for him, unlike the medium or high level student.

Lack of confidence

The instances of lack of confidence as is visible below came from only the low and medium level students:

“Because of the problems with use of English language, I lost confidence in writing. Writing was very hard as I imagined. Since I am not a good talker, I had troubles with expressing my thoughts in a proper way. This was all bad thing during this assignment.” -M5

“I realize that my writing skill learned in China cannot work in my English articles at all. Plus my vocabulary is not enough and English grammar is totally different from Chinese grammar, it is hard for me to write a good English article. When mistakes were pointed out in my first draft, I felt so upset. So when I talked to our teacher, I said ‘I afraid I can’t.’ But S (instructor) said ‘don’t say you can’t, you can,’ which made me feel encouraging.” -L2

“I felt not good, I tried my best to write it but it was so bad. Although I was so unhappy, I knew it was my problem. I should write my essay better. I came back home to rewrite my essay.” -L5

The last excerpt by student L5 is the most moving as we see the lack of confidence evident through his words. Even student L2 shares similar sentiments. However, since this is the first assignment L2 is talking about, it is possible that she was expecting more positive feedback than what she had gotten. This could explain her lack of confidence. Student M5 further brings in the component of speech, which he holds responsible for being unable to express himself clearly. These three excerpts depict the fears of these students which they were able to express in their journals.

Writing as fun

On a more positive note, there were excerpts that referred to writing as fun.

“In my time forging my personal narrative, I often found that the words flow out of me almost naturally. Overall, I found my personal narrative assignment to be emotionally-charged as it was primarily fueled by my feelings and opinions I desperately wanted heard. Talking about video games, a dim light in me instantaneously
flared up and I wrote with passion practically oozing out of my fingertips. Possessing a noble cause to battle the foul misconceptions of video games, I saw it as an eager opportunity and whole-heartedly seized it. To me, it was an essay that was already written out in my mind, right from the very start… to say I didn’t find this assignment enjoyable would be a lie. In the past, I have had many writing adventures due to free-for-all topics and had secretly found joy in these assignments.” –H1

“The good thing is that I felt that writing is interesting. I don’t have perfect grammar skills or writing skills, but when I wrote about personal narrative assignment, I felt that writing was fun. I just need to practice more on grammar and use of English language.” –M5

What is significant to notice here is that writing is fun was found in journal posts by high and medium level students. That is, low level students did not necessarily mention writing as fun. However, some low students always expressed hope that they would do better in their writing.

Comment codes

Following are some excerpts and examples of the comment codes. Notice the types of interaction between participants in the comments:

Empathy

L4 → H5: “I have the same problem of you, so I can clearly understand your feeling. The part of personal experience must be difficult if you don’t have enough information about it. And also, grammar problem also make me confused.”

M5 → L4: “As an international student, I understand that writing is hard, especially with second language. I understand how hard to organize thoughts and choose the correct words. I’m from Korea. Korean education system is similar to Chicness in some ways. In Korea, it is more important to write in correct grammar too. They don’t teach how to organize thoughts and write an essay.”

H1 → L1: “What you’re feeling is all too relatable to me. It reminds me of a time when I was still in a mandarin-orientated elementary school. Back then, my Mandarin skills were just awful and I remember dreading writing Mandarin essays. It was a challenge I never saw through because once I entered high school, I never took mandarin classes ever again. Fortunately for you, English writing is something that will stay with you in your time studying in America…”

Reading the above excerpts is absolutely heart-warming. This is where we begin to see the development of rapport between students. From the above, we can see a high student relate all too well with a low level student. Similarly, a medium, as well as, a low level student empathizes with their fellow classmates, too.

Encouragement

M1 → H4: “It seems you had everything very well plan since day one. The topic you chose is very interesting, and I can totally familiarize with it…I can foresee a very in-depth and captivating research paper to be written out of the previous explanation you wrote. Good Luck with everything!”

L2 → M1: “We all make a great progress in writing with the help of S and every classmate… From your journal, I can see you have a good time in this class, it’s nice. And confidence is a great thing. So keeping your confidence, you must will be better.”

H3 → M2: “…it’s good that you could find help to your problems and come up with a solution. It’s awesome that you could change your writing style to suite the requirement and make your narrative the best it could be. It takes great courage to write different from your style of writing, and I think you’ve been able to do that well as well as do it confidently. I wish you the best for the next few papers.”

As we can see in some of the encouragement codes above, students are seen providing support and also wishing well to their group members.
Compliments

H2→ L5: “First of all I would like to praise on your persistence, for not giving up writing a research paper in 4 hours despite all the challenges you faced. It is true that when you believe in yourself that you can do something, you eventually will find your way to do it. I love how you seek help from your friends and consistently try to improve your English writing and speaking skills.”

L5→ H2: “And I think your presentation was really good in the class. I did the presentation after you and I felt a little tense, because you did it so good. In overall your English are very good. I hope we can be friends and you can tell me how you study English well.”

L2 → H4: “Obviously, your arrangement for your paper is clearer than me. And there are so many materials around us which can help your paper. And it’s awesome that you had so many things to express (more than 1200 words).”

Here we see how fellow classmates are able to recognize the plus points of each other and compliment it. Once again, there is a shared sense of unity and understanding that the students are all in this together and hence support each other.

Advice/Suggestions

M5→ H5: “…if you try to describe little more details, you will get used to it. Also, you have to use a lot of words to describe, so you won’t have a problem with word limit. It is good that you satisfied your final draft. Also it is great that you figured all your problems out. If you work on describing details, you will find yourself writing the essay is easy.”

L2→ H4: “The way I think can solve this problems for you is to only write the greatest points, and write them more accurate. And I also wrote many words were too philosophical to express my paper in right way. It’s hard to avoid, right^.^ Hope our writing skills will be better.”

H2→ M4: “As for words usage, I recommend learning new words from synonyms because not only you can enhance your vocabularies, you can also avoid repetition especially when you have to refer to the same subject over and over again.”

The advice/suggestions code was the most interesting out of the comment codes. This is because ideally, one would hypothesize that students with higher proficiency level in a language would give more advice about that language. This is true in the last instance, above. However, it is interesting that both a medium and low level student give advice to high level students; whereas, there is only one instance of a high level student giving feedback to a medium level student.

Discussion

Through the above results a few points were visible when it came to the journal entries. The highest concern under journal entries was the challenge of dealing with linguistic errors, whether it was mechanics, or semantic features. This was followed by process/strategies that were adopted to cope with all challenges faced during the assignments. Accomplishments ranked third, and it is not surprising to note that while there were 11 accomplishment tokens featuring in the PN journal, only two accomplishment tokens featured in the LR journal. The personal narrative assignment, was the first assignment of the course. It was comparatively easier, since it required students to present a personal motivation for their choice of research topic. The LR journal posts, in contrast, depicted a stronger set of challenges for the students in the assignment. While, the high level students showed that they had language concerns, they also had trouble writing within the word limit. Nevertheless, they were the ones who enjoyed writing. Medium level students, on the other hand, mentioned having to deal with L1/L2 translation errors that affected their writing. However, they expressed a sense of accomplishment in terms of freedom of expression and language use. Low level students, showcased the highest instances of linguistic challenges that took up a major part of their journal posts. Nonetheless, it was these low students who adequately showed hope in wanting to do better in the class.
With regard to the comment entries, empathy and encouragement codes were the highest, followed by compliments, and a few instances of advice. What was interesting about the comments was that the medium level students tended to give more advice to low or high level students, instead of high students taking the lead. Could this be due to the possibility that medium level students have perhaps experienced both high and low levels of proficiency in their writing, and thus also experienced a variety of feedback? Some low level students were also found to be giving advice to high level students, which was interesting. Perhaps, this was because they probably receive the highest amount of feedback in their writing and thus have a higher feedback meta-knowledge to share? The above are my assumptions of course, and hence are difficult to conclude on, unless an interview can be conducted to corroborate my assumptions. Nevertheless, high and low level students mostly sympathized, gave encouragement, and compliments in their comments to others. The most important aspect to take away from this analysis of the commenting feature is how all the students managed to create a rapport system among themselves through this communication. Although they barely talked to each other in class, this online platform provided them an avenue to freely communicate with their fellow classmates and realize that they are not alone when it comes to composing in English in the classroom.

**Conclusion**

Although this study employed a small sample size, it provided the instructor a fresh perspective of where the students are in their writing. First, it allowed the instructor to develop and modify materials for future assignments. Second, it created bonds of trust and understanding between all participants, including the teacher spectator. More importantly, this research can be developed further to extend the findings of the current study. First, in future follow-ups of the study, the list of assignments and their corresponding journal entries can be expanded to include the interview and the argumentative paper assignments. This will provide a larger cohesive exposure to student challenges within the classroom, and would be highly beneficial, pedagogically. Second, the instructor can change the variables of viewership of the blog. The current study confined the viewers to only class participants and the teacher. This ensured a safe space for student expression, as they were not worried about public viewers and their opinion. Changing this variable can have positive effects on student confidence, since it would expose them to a wider audience and heighten their sense of ownership. Third, in this study, the instructor was a passive participant and made no comments on student exchanges and interaction. As long as students made timely journal entries, they would get a full grade on the assignment. This was done on purpose to ensure a low stress environment for student expression. However, opening the possibility of teacher involvement in the future, will provide the scope for formative assessment during the writing experience, as opposed to only summative assessment witnessed in this study. Maintaining the class size and number of assignments is however, important, since this will ensure that students are not over-burdened, and the instructor is able to sufficiently cater to the needs of all his/her students. Some points that however need to be considered are: students need to be aware of what journaling is before such a task is undertaken in the classroom. In other words, students should be trained in journal writing so as to effectively use it for writer-growth. Additionally, teachers must consider motivational factors of students’ journaling process and consciously work to increase their motivation to reflect. In this way, journaling can definitely be helpful and effective in not only writing, but personal growth, as well.

**References**


Appendix A: Journal prompt

These journal posts should discuss your reflection on how your writing has progressed during the course of the recently finished assignment. Discuss the challenges, inspirations, annoyance and anything and everything you felt with regard to improving that specific assignment. If there was a specific incident that helped or affected your writing, write about it too. You have to have at least 300 words for these journal entries, but feel free to write more if you would like!

Appendix B: Comment prompt

You should know who your two group members are. Respond to both your group members’ reflection by relating to it in any way you can. Your response should be at least 100 words, but feel free to write more if you would like!
The Effect of Computer Game-Based Learning on FL Vocabulary Transferability

Stephan J. Franciosi
Osaka University, Kyoto City, Japan // steve.franciosi@gmail.com

(Submitted November 20, 2015; Revised February 24, 2016; Accepted March 16, 2016)

ABSTRACT

In theory, computer game-based learning can support several vocabulary learning affordances that have been identified in the foreign language learning research. In the observable evidence, learning with computer games has been shown to improve performance on vocabulary recall tests. However, while simple recall can be a sign of learning, observation of skill application in communication is a better indicator of skill mastery. Further, observing this use in separate communicative contexts could constitute evidence of transferability of skills. Hence, this paper presents the results of two investigations of learning outcomes in EFL classes at a Japanese university using computer game-based lessons. The first study was a quasi-experiment comparing use of targeted words in a writing task between a group of students who participated in a computer game-based lesson, and a group of students who did not. The second study was a cross-sectional analysis comparing use of targeted vocabulary in a writing task with amount of participation in computer game-based lessons. The results suggest that computer game-based approaches to foreign language education in real-world classrooms can improve transferability of learned vocabulary.

Keywords

Computer game-based learning, Foreign language learning, Vocabulary learning, Transferability

Introduction

This paper reports on two studies to examine the effectiveness of computer game-based approaches in foreign language learning. A game-based approach follows mainstream foreign language education models that prescribe a “meaning-focused” activity wherein learners apply the target language to perform a task, supported by “form-focused” enabling tasks wherein learners learn and/or practice linguistic form, and followed by a post-activity phase for reflection and knowledge construction (Ellis, 2003). “Game-based” here simply denotes the use of a computer simulation game as the meaning-focused activity.

Computer games could theoretically provide several language learning affordances described in the literature on second language acquisition, and a growing body of empirical data on game-based approaches showing improved retention of learned words supports this notion (e.g., Ranalli, 2008).

This paper seeks to address two issues with the extant literature. Firstly, studies showing positive impacts on vocabulary acquisition as a result of using games rely heavily on data elicited with contrived vocabulary testing instruments as indicators of word retention, but many SLA researchers consider language data produced spontaneously in communicative acts as a more direct indicator of foreign language ability (Norris & Ortega, 2003). Secondly, simple retention may not be the only benefit of providing these learning affordances. Klimesch’s (1994) Connectivity Theory predicts that the same learning affordances that improve performance on recall tests can promote transferability of learned words. It is important to attain evidence that instructional techniques can enhance transferability if the goal of foreign language education is to foster the skills needed to use the language in various communicative contexts.

In view of the preceding, the purpose of this paper is to report the results of two studies of vocabulary learning outcomes from EFL lessons employing Energy City, an online simulation game, as core material in the meaning-focused activity. The first study was a quasi-experiment comparing spontaneous usage of targeted vocabulary in a writing task between a group of learners who learned the words in a drill application and subsequently used the words in a game-based lesson, and a group of learners who learned vocabulary with the drill application alone. The second study was a cross-sectional analysis of student work comparing amount of game-based learning with Energy City to spontaneous usage of targeted vocabulary in a separate writing task.
Review of literature

Vocabulary and game-based language learning affordances

Language use

The literature suggests that the use of targeted vocabulary in the proactive performance of some type of task improves recall of those words. Output-oriented activities such as writing facilitate word retention more than reading (Mármol, & Sánchez-Lafuente, 2013; Jahangard & Movassagh, 2011), and sentence scramble exercises are more effective than gap-fill (Haratmeh, 2012). Input-oriented tasks also appear to positively influence vocabulary acquisition when learners are required to perform an action based on the meaning of the input (Shintani, 2012). Overall, these studies support the notion that using language in the performance of tasks requires more “depth of processing,” proposed by Craik and Lockhart (1972). Deeper processing implies extensive cognitive processing such as that involved in searching and decision making, and is posited to facilitate learning and retention.

The proactive role of the player in games could facilitate deep processing of concepts represented by targeted words. Agency is an integral component of games and gameplay (e.g., Gee, 2005; Salen & Zimmerman, 2004). The ability to influence the game system presumably requires players to consider, before taking action, how various factors might influence future game states, and the nature of possible outcomes. This premeditation could evoke the deep processing proposed by Craik and Lockhart (1972). Thus, agency may positively influence vocabulary learning and memory.

A further way that gameplay could influence the use of learned vocabulary is through spontaneous and/or informal communication between game players outside of the game itself. Games are known to engender communities of players that engage in considerable knowledge sharing (Nardi, 2010). Ryu (2013) analyzed discourse generated by an online game community and concluded that this aspect of “game culture” provides ideal opportunities for foreign language learning and practice. Additionally, gameplay has been shown to lower inhibitions and stimulate meaningful communication among language learners (Reinders & Wattana, 2011; Wu, Chen, & Huang, 2014). Thus, games seem capable of facilitating natural discourse among players.

In addition to spontaneous and informal communication, game-based learning can also involve formalized discourse. Most game- and simulation-based learning scholars are strong advocates of including debriefing as an integral component of instruction (Crookall, 2010). The insistence that debriefing be included is based on Kolb’s (1984) idea that learning optimally comprises experiences (such as gameplay) interspersed with periods of reflection on those experiences. For example, Kriz (2008) prescribes a semi-structured discussion approach to guiding post-game discussion, and Oertig (2008) reports on the use of a forum to allow learners to share written debriefing reports. While the objective of debriefing for these authors is presumably to learn about non-linguistic concepts in a game or simulation, the process of debriefing itself obviously involves communicative use of language. Whether using language to communicate about a game is spontaneous or formalized, spoken or written, the literature reviewed here indicates that doing so has a beneficial impact on vocabulary learning and memory.

Association with non-linguistic information

Association of linguistic and non-linguistic information has also been found to facilitate vocabulary learning and memory. Such information includes audiovisual cues (Chuang & Ku, 2011), particularly dynamic images as in video (Al-Seghayer, 2001; Hsu, 2014), as well as iconic physical movements such as gestures (Macedonia & Klimesch, 2014). The presence of non-linguistic information at the time of learning seems to involve activation of areas of the brain associated with non-linguistic as well as linguistic processing, and this increased activity correlates with superior performance on recall tests for novel words (Takashima, Bakker, van Hell, Janzen, & McQueen, 2014). Thus, diversity of information available to a learner at the time of learning may have a positive impact on vocabulary retention.

Computer games can present non-linguistic (audiovisual) information together with targeted linguistic constructs. Sophisticated graphics and audio are commonly considered fundamental components of most computer games (Rollings & Adams, 2003; Rouse, 2005). Although the literature cited above indicates that associating words with images and sounds enhances recall, the studies on the effects of multi-media game environments on learning demonstrate mixed results. For example, “low-immersive” environments appear to
assist learning more than “high-immersive” ones (Schrader & Bastiaens, 2012), but 3D virtual world environments have been reported to be superior to standard paper textbooks (Barab et al., 2009). However, Neville and Shelton (2009) point out that sensory immersion can and should be adjusted to accord with specific learners and learning goals. In sum, the literature indicates that non-linguistic sensory information in games may support vocabulary learning provided that it is not overwhelming.

Association with interesting scenarios

Learners recall words more readily when they judge the concepts represented by the words as success-relevant to a compelling problem-solving scenario. Novel words for items relevant to a survival scenario in the wilderness are better recalled than novel words related to moving a residence in the city (Nairne, Thompson, & Pandeirada, 2007), and this “survival processing effect” may result from a sense of urgency that motivates learners to brainstorm multiple roles for these items (Toyota & Kikuchi, 2005). Interestingly, Soderstrom and McCabe (2011) discovered that learners are more likely to recall words relevant to a survival scenario involving zombies compared to one involving natural predators or human attackers, suggesting that strong affective factors (e.g., fear, excitement) related to an imagined scenario can play a role in facilitating memory and recall. Overall, this research indicates that level of interest in a scenario can influence vocabulary learning.

One of the hallmarks of many games, particularly those commonly classified as belonging to the role-playing, simulation or strategic genres, is the creation of compelling stories comprised of interrelated situations. Game designers strive to give players an obstacle to overcome, a problem to solve, or a decision to be made in each situation (Koster, 2005). Ideally, these challenges motivate learners to search for a variety of problem-solving means. Further, strong emotions may also result when the nature of the contrived situations matches the interests of the players, but not every challenge or scenario presented by a game is compelling for every learner (Schell, 2008). Therefore, games could facilitate vocabulary learning depending on the extent to which learners are interested in the subject matter represented by the game.

Evidence of language learning in games

Several studies have uncovered evidence of efficacy regarding vocabulary learning in computer games and simulations. These studies have used a variety of measurement protocols, such as vocabulary recognition, multiple choice, matching, cloze tests (Bakar & Nosratirad, 2013; Berns, Gonzalez-Pardo, & Camacho, 2013; Chen, 2014; Milton, Jonsen, Hirst, & Lindenburn, 2012; Ranalli, 2008), and a Vocabulary Knowledge Scale in which participants self-report on their mastery of a word (Smith et al., 2013). Thus, the effectiveness of computer games in vocabulary learning has been demonstrated.

The aforementioned studies largely relied on decontextualized measures of lexical knowledge to make their conclusions, with the exception of Milton et al. (2012) who also analyzed discourse generated among learners during gameplay. Such instruments are useful for indicating retention, but the extent to which they can be used to demonstrate transferability is more tenuous. Even in the case of Milton et al. (2012), the implications for transferability are unclear because the language analyzed was produced during the learning activity studied. Furthermore, while data obtained with elicitation devices such as quizzes and tests can be useful indicators of second language development (Ellis, 2008), spontaneous use of language in a communicative act is usually considered a more convincing sign of actual linguistic proficiency (Norris & Ortega, 2003). Furthermore, in Hsu’s (2014) study of audiovisual aids for writing, the author considered such use of vocabulary as more indicative of mastery than passive word recognition because the choice of words to produce was left to the learners who were free to, “avoid words that they considered problematic or were unsure of,” (p. 62). The present author further assumes that “mastery” of a skill entails the ability to use it productively in a variety of contexts, and thus use of vocabulary from a game in a separate task would suggest that game-based learning improves transferability of targeted words.

Learning and memory theory

In Connectivity Theory, Klimesch (1994) posits the metaphor of interconnected nodes to describe knowledge structure. Nodes represent a variety of information such as motor-sensory, emotional, and linguistic. Nodes become associated with each other in learning, and can have connections with multiple nodes forming a network. Klimesch posits that increased complexity of a network in terms of number and variety of connections increases
the retention and accessibility of the information contained therein. This notion is supported by Macedonia and Klimesch (2014), cited above, who found evidence that associating novel words with physical movements improves long-term retention. Klimesch’s conceptualization of learning and memory is compatible with the observed evidence cited above. Deep processing, association with audiovisual information, and compelling problem-solving scenarios that evoke strong emotions would all serve to increase the number of interconnections and diversity of nodes in a network.

A further logical extension of the connectivity account of vocabulary knowledge is that the more interconnected a target language word becomes, the more transferable it is (i.e., the more readily it becomes usable in separate contexts). The reason is that nodes forming one network may belong to multiple separate networks which could be relevant for a variety of communicative situations. Thus, the more complex the network to which a target language word belongs, the more likely it would be usable in a situation outside of the environment in which it was learned. Learners using words spontaneously in a communicative act outside of the context in which they were learned would support the Connectivity account of knowledge structure, and constitute evidence of transferability.

Study 1

Method

The first study was a quasi-experiment comparing quantitative use of targeted vocabulary in a writing task between a group of learners who learned vocabulary with a drill application and subsequently used the words in a game-based lesson, and a group of learners who learned vocabulary with the drill application only.

Participants

This study employed a sample of convenience using students enrolled in four EFL courses taught by the author at a Japanese university. All classes were required general education courses for first and second year students. The nature of participation in the study was contingent on internet availability. One of the four classrooms was equipped with wireless network antennae allowing internet access from student laptop devices, so the author determined to employ game-based lessons using an online computer game in this class (n = 23), and compare learning outcomes with the remaining three classes (n = 61).

Instructional materials

Energy City

The online game simulates energy production and conservation strategies in urban settings (Figure 1).
This activity is considered “meaning-focused” in that while linguistic labels and explanations for important features are provided, operation of the simulation requires attention to be paid primarily to non-linguistic factors, such as the relative costs and benefits of various energy production technologies. The object of the game is to plan and implement a virtual city’s energy portfolio in order to maintain a sufficient supply of electricity over the course of twenty turns without overly imparting damage to the environment or depleting the city budget. The author chose this game because he assumed that the topic would be a compelling one for Japanese students given the increase in public discourse on the topic following the Fukushima nuclear disaster (Aldrich, 2013).

**Debriefing session**

The author included a debriefing session as an integral component of game and simulation based instruction according to best practices in game-based learning (Crookall, 2010). Following the tenets of Experiential Learning Theory (Kolb, 1984), the purpose of debriefing components is to facilitate critical reflective thought following experience in meaning-focused activities, thereby supporting new knowledge synthesis. The session was a semi-structured discussion generally following Kriz’s (2008) recommendations. The discussion followed five questions designed to elicit commentary on:

- affective state as a result of the experience
- memory of details
- new information garnered
- possible real-life applications of that information
- future plans for continued learning

**Enabling task**

Best practices in foreign language instruction prescribe the use of a form-focused enabling task to support language learning and accuracy (Ellis, 2003). Form-focused activities are typically characterized by repeated application of a narrow skill set. Hence, the present author made a study set in the online vocabulary learning application, Quizlet for 33 keywords from Energy City which, in the author’s judgment, were likely unfamiliar to most of the learners. Quizlet supports six learning activities designed to train a variety of form-focused vocabulary skills, such as word recognition and spelling. The core mechanic for each of these activities is matching, and because all of the participants were either native or fluent speakers of Japanese, the author configured the study set so that learners would be required to match targeted English vocabulary items with their Japanese counterparts.

**Data collection tools**

**Vocabulary pretest and posttest**

Because this study did not involve random selection, the author used a vocabulary recall test to establish a baseline for preexisting knowledge of the words in the vocabulary list as a means of improving validity. Additionally, in order to account for variation in mastery of the vocabulary items through use of Quizlet, the author also developed a post test. Items prompted respondents to write an English word by showing the Japanese counterpart. Both instruments tested knowledge of all 33 words in the Quizlet study set in separate random orders. The instruments were deployed on a Google form and scored with a search algorithm that counted unique occurrences of targeted words in the text.

**Writing task**

The author collected data on word use using a short writing task. Participants responded in two paragraphs to two writing prompts. The first paragraph was designed to be expository, wherein participants were asked to describe the current state of energy supply and use in Japan in approximately 50 words. The second was designed to be persuasive, wherein participants were asked to describe how they would change the status quo if they could or would in approximately 100 words. This task was deployed on a paper-based instrument, and later inputted into a Google sheet for scoring with a search algorithm that counted unique occurrences of targeted words in the text.
Procedures

Instructional treatment followed class schedules for Japanese universities, which hold classes once a week for 90 minutes. Treatment began in the second week of the semester after enrollment had closed.

Game-based learning group

In the initial treatment session, the students were pretested in class on their mobile devices (ownership of which was determined to be ubiquitous) and assigned the Quizlet study set as homework. They were informed that they would be tested again on the vocabulary items the following week. In the second session, the students were first post-tested for mastery of the vocabulary list, after which Energy City was introduced and demonstrated on a large screen for approximately ten minutes. Students were then asked to play the game in small groups of three to four members on their own laptop devices for up to 60 minutes. The author provided students with paper copies of the debriefing report questions, informed them that these would be discussed in class the following week, and invited them to prepare their responses. In the final session, students debriefed in small groups and the writing task was assigned to be completed in class without dictionaries.

Control group

In the initial treatment session, the students were pretested in class and assigned the Quizlet study set as homework so that the quasi-experimental comparison would focus on the game/debriefing method. Students were informed that they would be tested again on the vocabulary items the following week. In the second session, the students were posttested in class and the writing task was assigned to be completed in class without dictionaries.

Analyses and results

Pre- and posttests and the writing task scores were inputted into NCSS 9 for analysis. An ANCOVA was conducted to determine any difference in the amount of targeted words used in the writing task between the game-based lesson group and control group, while accounting for preexisting knowledge of the words by using the pretest scores as covariate. There was a significant effect of game-based learning on the number of vocabulary list words used in the writing task after controlling for pretest scores; $F(1, 81) = 10.41, p = .002$. Omega squared was selected as benchmark for effect size because it is considered the least biased indicator when dealing with relatively small sample sizes, such as in the present study. Thus, the effect size was determined and interpreted as “medium” ($\omega^2 = 0.1$) according to Field’s (2013) guidelines (p. 474) by calculating omega squared according to the following equation:

$$\omega^2 = \frac{SS_{\text{effect}} - (df_{\text{effect}})(MS_{\text{Error}})}{SS_{\text{Total}} + MS_{\text{Error}}} = \frac{40.5 - (1)(3.9)}{360.7 + 3.9} = 0.1$$

A second ANCOVA was conducted to determine any difference in the amount of targeted word usage between groups while accounting for variation in mastery of the vocabulary with Quizlet by setting the posttest scores as covariate. Significance was detected in this test as well; $F(1, 81) = 10.13, p = .002$. The effect size was determined and interpreted as “medium” ($\omega^2 = 0.09$) according to Field’s (2013) guidelines (p. 474) by calculating omega squared according to the following equation:

$$\omega^2 = \frac{SS_{\text{effect}} - (df_{\text{effect}})(MS_{\text{Error}})}{SS_{\text{Total}} + MS_{\text{Error}}} = \frac{37.8 - (1)(3.7)}{360.7 + 3.7} = 0.09$$

As shown in Figure 2, students who experienced the game-based lesson after learning with Quizlet used words from the study set an average of approximately four times in the writing task, with 50% of them using between two and five of the words. By contrast, students that only learned with Quizlet used an average of approximately two of the words, with over 75% of them using fewer than four.
Study 2

Method

The second study was a cross-sectional analysis of student work comparing amount of game-based instruction with *Energy City* to spontaneous usage of targeted vocabulary in a separate writing task. This investigation relied on natural variation in student completion of class work to indicate causal relationships.

Participants

The study employed a sample of convenience from students enrolled in four EFL courses taught by the author at a Japanese university (n = 143). The classes were the same as in Study 1, but held in the following semester with a different group of students.

Instructional materials

This study employed the same simulation game and enabling task used in Study 1. One important difference in use was that instead of playing only one of the six city scenarios available in *Energy City*, Study 2 utilized all six. Also, the questions for the debriefing report were deployed online using a Google form.

Data collection tools

This study employed the same writing task and scoring algorithm as described in Study 1. Additionally, the present author asked the students to submit a score from the Quizlet test activity. The test activity of Quizlet randomly generates a 20-item test employing written, matching, multiple choice and true/false questions. Finally, the author assumed that the number of debriefing reports submitted online was an indicator of the amount of participation in game-based learning (i.e., the assumption is that no student submitted a report without first playing the game).

Treatment and data collection procedures

The Quizlet study set was first assigned as homework. Students were instructed to use the activities for one week and submit their best score on the test activity. In the following session, *Energy City* was demonstrated in class and assigned as homework. Students were instructed to play each city scenario individually and submit a
debriefing report online. The timing of completion was left up to the students, but they were given six weeks to complete all six scenarios. A week after the deadline, students were assigned the writing task as homework.

**Analyses and results**

*Quizlet* test scores, the number of debriefing reports submitted, and scoring results of the writing task were inputted into NCSS 9 for analysis. A correlation coefficient was calculated to assess the relationship between the amount of participation in the game-based lessons ($M = 5$, $SD = 2$) and the unique occurrences of vocabulary list words in the writing task ($M = 4.9$, $SD = 2.8$). There was a small positive correlation between the two variables, $r = 0.29$, $n = 143$, $p = .0004$. As shown in Figure 3, the more the students completed the game-based lessons, the more likely they were to use keywords in the writing task. The data indicate that participation in the game-based lessons accounted for 8% of the variation in targeted word use in the writing task ($R^2 = 0.08$). The effect size of the correlation ($r$) is “small” according Cohen’s (1988) benchmarks.

Mastery of vocabulary by virtue of *Quizlet* may also influence keyword use in the writing task. Thus, a correlation coefficient was calculated to assess the relationship between the test score ($M = 86.5$, $SD = 34$) and unique occurrences of vocabulary list words in the writing task. There was no correlation between the two variables, $r = 0.13$, $n = 143$, $p = .11$. Mastery of vocabulary in *Quizlet* had no apparent influence on targeted word use in this study.

![Figure 3. Scatter plot of submitted reports and targeted words](image)

**Discussion**

Overall, the results of both studies suggest that instructional approaches using simulation games can support vocabulary transferability in second language learners. Participation in the lessons using *Energy City* as core material is related in both studies to the number of targeted vocabulary words used in the writing task. The writing task comprises a communicative context that is distinct from learning vocabulary with *Quizlet*, playing the game or participating in a debriefing session. This result is predicted by Connectivity Theory (Klimesch, 1994), which posits that the learning affordances provided in the lessons would lead to greater interconnection between the targeted words and a variety of other information, and that this in turn would increase accessibility to the words allowing learners to use them more readily in separate contexts. Transferability is important to acknowledge because one of the primary purposes of foreign language teaching is to impart communicative skills which can be used outside of the learning context.

Notably, both studies were conducted in actual educational contexts. Several of the studies supporting computer game-based language learning that have occurred in laboratory conditions are useful for offering guidance, but pedagogy is more reliably informed when models are field-tested in real-world classrooms. The present paper demonstrates the effectiveness of game-based learning when not every variable can be controlled, and within the constraints and requirements of genuine courses of study.
Also noteworthy was the difference in effect size between the two studies ("medium" in Study 1, and "small" in Study 2). Three design differences between the investigations could account for this. First, in Study 1 the writing task was carried out immediately following the debriefing session, while in Study 2 these tasks were carried out individually as homework. Therefore, learners may have felt a social pressure to apply more effort during the task in the former, thereby improving their attentiveness and learning. Finally, while the writing task was conducted under instructor supervision without the use of dictionaries in Study 1, learners were left to complete the task on their own in Study 2 without the use of dictionaries being explicitly prohibited. Therefore, it is possible that the use of dictionaries acted to diminish the skill disparities between learners who had participated in game-based lessons and those who had not. Unfortunately, based on the present data, it is not possible to determine which if any of these factors played a significant role in creating the discrepancy in effect size.

The results of the studies presented in this paper include some further limitations that must be noted. First, based on data reported in this study, the author makes no claim that computer games are superior in terms of learning outcomes compared to other commonly used meaning-focused activities. Non-game activities in task based approaches include, for example, presentations, discussions, debates, or project work. Such lessons may very well provide all of the learning affordances described herein which are attributable to certain computer games, and the present data does not demonstrate otherwise. Second, no claims can be made regarding the respective degrees of effect among the lesson components. In other words, the game-based lesson must be treated as a single factor encompassing the enabling task (Quizlet), the game itself as well as the debriefing reports. The data can only suggest that this integrated lesson is superior to the enabling task alone at supporting transferability of vocabulary. Yet, such enabling tasks are likely useful for increasing scores on other measures of foreign language proficiency, such as the recall tasks used for pre- and posttesting in Study 1. Finally, although this particular game, Energy City, appeared to be effective for this particular group of learners, Japanese university students, the author does not suggest that it would necessarily work well with other demographics. Interest in a topic portrayed in a computer game depends largely on the individual user, as pointed out by Schell (2008), and whether or not the scenario is compelling for learners likely influences outcomes (Soderstrom & McCabe, 2011). Thus, the conclusion based on the present data must be limited to the claim that computer games can be used successfully to support transferability provided that best practices in applied linguistic, game and simulation methodology are followed (enabling and debriefing tasks are included), and that the subject matter matches learner interest.

Yet, this conclusion, although limited, is an important one. Educators who consider using games in their classes may do so in recognition of a positive motivational effect on students, which has been demonstrated many times in the literature (Ballou, 2009; Baltra, 1990; Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Liu & Chu, 2010; Roubstova, 2010; Suh, Kim, & Kim, 2010; Tao, Cheng, & Sun, 2009; Woo, 2014). At the same time, they may hesitate to use games, or marginalize their use, because of a perceived lack of effectiveness in terms of learning outcomes. The results reported in this paper strongly suggest that using computer games as a meaning-focused activity has a beneficial impact on language learning if the goal is to engender the ability to use learned skills in different communicative situations. Therefore, concern over learning outcomes alone should not dissuade educators from employing games.

**Conclusion**

This paper reported the results of two investigations on the learning outcomes of computer game-based lessons in two learning programs. The first study compared use of targeted words in a writing task between a group of students who learned the words with Quizlet and subsequently used them in a computer game-based lesson, and a group of students who only learned the words with Quizlet. The second study compared the relationship between participation in a game-based lesson and use of targeted words in a writing task. The author concludes that computer games can be used effectively as meaning-focused activities in foreign language education.

**References**


Surveying In-Service Teachers’ Beliefs about Game-Based Learning and Perceptions of Technological Pedagogical and Content Knowledge of Games

Chung-Yuan Hsu¹, Meng-Jung Tsai², Yu-Hsuan Chang² and Jyh-Chong Liang²

¹Department of Child Care, National Pingtung University of Science and Technology, Pingtung, Taiwan // ²Graduate Institute of Digital Learning and Education, Taipei, Taiwan // mjtsai99@mail.ntust.edu.tw // M10111017@mail.ntust.edu.tw // aljc@mail.ntust.edu.tw

*Corresponding author

(Submitted December 3, 2015; Accepted January 19, 2016)

ABSTRACT

Using the Game-based-learning Teaching Belief Scale (GTBS) and the Technological Pedagogical Content Knowledge—Games questionnaire (TPACK-G), this study investigated 316 Taiwanese in-service teachers’ teaching beliefs about game-based learning and their perceptions of game-based pedagogical content knowledge (GPCK). Both t-tests and ANOVA analyses were used to examine the differences due to their demographic backgrounds. Correlation and regression analyses were used to examine the predicting variables of the teachers’ GPCK. The results showed that, among the factors of GTBS and TPACK-G, GPCK plays the most critical role in predicting GPCK. In addition, the elementary school teachers had stronger Belief, Confidence and Motivation for using game-based teaching approaches, and had higher GPCK than middle school teachers. A gender difference only existed in GK. Younger teachers had better GK and GCK than older teachers. Compared with experienced teachers, novice teachers tended to believe that digital games can assist learning and instruction, and perceived higher self-efficacy in their TPACK-G.

Keywords

Game-based learning, TPACK, Games, Teacher, Teaching beliefs

Introduction

Digital games (hereafter named games) have been gaining tremendous interest as tools for teaching and learning in recent years. The merits of game-based learning (GBL) include supporting effective learning (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013), enhancing higher order thinking (Sánchez & Olivares, 2011; Yang, 2015), increasing problem-solving skills (Akcayoglu & Koehler, 2014; van de Sande, Segers, & Verhoeven, 2015), and promoting engagement (Annetta, Minogue, Holmes, & Cheng, 2009; Hsieh, Lin, & Hou, 2015). Despite the potential of games as tools for teaching and learning, teachers, as indicated by Klopfer, Osterweil, and Salen (2009), could possibly face barriers to adopting games in class. For instance, they might have little experience of integrating games into the classroom, and they might have problems engaging students due to insufficient understanding of the variety of games available. Most importantly, successful models of integrating pedagogy, content, and games into a range of curricular experiences are limited, which may reduce teachers’ willingness to utilize games in class. While game-based learning is potentially useful for improving teaching and learning, finding a theoretical framework that helps to probe practitioners’ knowledge of teaching with games has become crucial.

A potentially helpful framework for evaluating and guiding the integration of technology into classroom teaching and learning is technological pedagogical content knowledge (TPACK) (Chai, Koh, Tsai, & Tan, 2011). Due to the popularity of technology in education, Mishra and Koehler (2006) extended Shulman’s (1986) pedagogical content knowledge (PCK) by adding technological knowledge (TK) as an integrated part of pedagogical knowledge (PK) and content knowledge (CK). Successful integration of technology into the classroom requires one to focus on the complex interplay of these three forms of knowledge (Mishra & Koehler, 2006). While the notion of TPACK has received attention and approval, many researchers (e.g., Hsu, Chai, Liang, & Tsai, 2013; Lee & Tsai, 2010) quickly found that the current TPACK studies tend to treat technology in a general manner. This may mean that the TPACK framework is unable to provide guidelines for teaching with a specific type of technology such as games. Thus, building on Mishra and Koehler’s (2006) TPACK, Hsu et al. (2013) proposed a framework of Technological Pedagogical Content Knowledge—Games (TPACK-G). The TPACK-G includes game knowledge (GK), game pedagogical knowledge (GPK), game content knowledge (GCK), and game pedagogical content knowledge (GPCK). GK is knowledge about the general usage of games; GPK is knowledge about using games to implement different teaching methods; GCK is knowledge about subject matter
representation with games; and lastly, GPCK is knowledge about using games to implement pedagogical strategies for teaching any subject matter content.

Many researchers and educators have indicated that teachers play the dominant role in using games in the formal educational context (Blackwell, Lauricella, Wartella, Robb, & Schomburg, 2013; Celik & Yesilyurt, 2013; Ketelhut & Schiffer, 2011). This is due to the fact that teachers’ teaching behaviors are influenced by their beliefs, confidence and motivations for teaching. For instance, teachers who believed that technology works best for instruction were found to be able to integrate technology into their teaching practices (Blackwell et al., 2013; Ertmer, Ottenbreit-Leftwich, Sadik, Sendurur, & Sendurur, 2012; Kordaki, 2013). Teachers with higher confidence in technology are likely to effectively and significantly succeed in technology-related tasks (Celik & Yesilyurt, 2013; Teo, 2009). Motivated teachers tend to integrate ICT into their classroom teaching as well as displaying continuous usage of technology in class (Sang, Valkce, van Braak, Tondeur, & Zhu, 2011). Thus, exploring the factors contributing to the individual differences in teachers’ teaching beliefs about using games for teaching has become an essential issue in the game-based learning research.

Individual personal factors such as age, gender, teaching experience and teaching levels usually play important roles in teachers’ teaching beliefs, confidence, commitments and motivations for teaching. Thus, the demographic factors such as gender, age, and teaching experience may also influence teachers’ technology use in class. Researchers (Erdogan & Sahin, 2010; Jang & Tsai, 2013) have reported that male teachers were more competent in their technology-related knowledge (e.g., TK and TPCK) than female teachers. Others (Lin, Tsai, Chai, & Lee, 2013) also indicated that the older the in-service science teachers, the less their confidence in embedding technology into their teaching. Similarly, in Lee and Tsai’s (2010) study on probing teachers’ self-efficacy for teaching on the Web, teachers with more teaching experience tended to have lower self-efficacy for using the Web for instruction. Lastly, in Taiwan, middle school teachers are recruited to teach the subject based on their undergraduate major, whereas elementary school teachers are generalists. This may result in different GTBS and TPACK perceptions according to teaching level, which has rarely been examined. Moreover, one issue which needs to be noted in Taiwan is the difference in the ICT acceptance and implementations between elementary schools and secondary schools. For example, in a recent ICT project initiated by the Ministry of Education (2015) in Taiwan to enhance future citizen’s literacy education, 80 percent of the projects were applied by elementary schools, whereas only 20 percent came from middle schools. These different participation rates may be related to the different motivations, willingness and attitudes toward using ICT of teachers at the different school levels. Therefore, it is important to explore if there are any differences in elementary and secondary school teachers’ teaching beliefs and self-efficacy for using new technology such as games for teaching.

In sum, researchers have indicated the influences of teachers’ demographic variables (e.g., age, gender, teaching experience) and affective variables (e.g., confidence, motivation) on their behaviors of integrating technology into education. However, little research has been specifically conducted to establish a similar link to teachers’ classroom uses of games, especially in terms of the teaching beliefs regarding games-based learning. The present study aimed to fill this gap. The purposes of the study were to answer these two questions:
- What are Taiwanese in-service teachers’ teaching beliefs about game-based learning (GTBS) and TPACK-G perceptions in terms of their demographics (teaching level, gender, age and teaching experience)?
- How do teacher demographics and GTBS (Beliefs, Confidence, and Motivation) and the TPACK-G constructs (GK, GPK, and GCK) predict in-service teachers’ GPCK?

Methodology

Participants

The participants of this study included 316 in-service teachers teaching in the northern, middle and southern parts of Taiwan. Within this sample, 176 were elementary school teachers (45 males and 131 females) and 140 were middle school teachers (36 males and 104 females). The average age was 38.80 (SD = 7.78) and the average teaching experience was 14.29 years (SD = 8.51). Regarding the educational background, 50.9 % of the teachers had a bachelor’s degree, and 49.1 % had a master’s degree. The teachers reported that their average game playing experience was 3.15 years (SD = 4.92). After completing the surveys, some participants were randomly selected for interview to probe their beliefs about game-based learning.
Instruments

The questionnaires, Game-based-learning Teaching Belief Scale (GTBS) and Technological Pedagogical Content Knowledge—Games (TPACK-G), were employed in this study. The GTBS was revised from Chang and Tsai’s (2014) previous work on assessing teachers’ teaching beliefs about game-based learning. The original version (Chang & Tsai, 2014) included 13 items categorized into the three subscales of Belief, Attitude and Motivation, with an overall reliability of 0.87. Through a further validation using a confirmatory factor analysis, four items with a factor loading of less than 0.50 and with many cross loadings were deleted. The final GTBS scale used in this study consists of the following three factors: Belief, Confidence, and Motivation. Descriptions of the three subscales are presented below:

- **Belief**: assessing teachers’ subjective and firm perspectives, values and trustworthiness of digital game-based learning, such as “I believe that digital games provide students opportunities to solve problems.”
- **Confidence**: measuring teachers’ experience and confidence in adopting digital games in teaching, such as “I have no idea how to integrate digital games into my curriculum” (scored in reverse).
- **Motivation**: assessing teachers’ commitment and willingness to adopt digital game-based learning for teaching in the future, such as “When I prepare my teaching plans, I will link my curriculum with digital games.”

The final GTBS scale consisted of 9 items which were categorized into 3 subscales. They were measured using a seven-point Likert scale, ranging from strongly disagree to strongly agree. In this study, the reliability coefficients were 0.83, 0.87, 0.86, respectively for the Belief, Confidence and Motivation subscales, and 0.87 for the overall scale. This suggests that this scale is highly reliable in terms of assessing teachers’ teaching beliefs and commitments to using GBL in teaching.

The TPACK-G survey was developed by Hsu et al. (2013) to explore teachers’ confidence in TPACK-G. Modified from Koh, Chai, and Tsai’s (2013) and Lee and Tsai’s (2010) survey items, this instrument includes the factors of game knowledge, game pedagogical knowledge, game content knowledge, and game pedagogical content knowledge. Descriptions of the factors are presented below.

- **Game knowledge (GK)**: measuring teachers’ confidence in using digital games, such as “I can learn digital games easily.”
- **Game pedagogical knowledge (GPK)**: exploring teachers’ confidence in using digital games to enhance students’ learning, such as “I know how to use characteristics of digital games to support teaching.”
- **Game content knowledge (GCK)**: assessing teachers’ confidence in using digital games to represent specific content, such as “I can identify whether the core concepts of the subject matter knowledge are displayed in the digital games.”
- **Game pedagogical content knowledge (GPCK)**: measuring teachers’ confidence in utilizing appropriate pedagogy and digital games to support students’ learning of specific content, such as “I can teach lessons that appropriately combine my teaching subject, digital games and teaching approaches.”

The TPACK-G survey includes 22 items presented with a seven-point Likert scale, ranging from strongly disagree to strongly agree. The Cronbach’s alpha coefficients respectively for these scales were 0.94, 0.95, 0.97 and 0.96. The overall alpha was 0.96, suggesting that these scales had high reliability for assessing teachers’ confidence in their TPACK-G.

Data analysis

A series of independent t-tests was conducted to examine any difference in the participants’ GTBS and TPACK-G in terms of their teaching level and gender. The participants’ ages and years of teaching experience were separately categorized into three levels. The former included the levels 21-30, 31-40, and over 40, while the latter included the levels 0-10, 11-20, and over 20. Analyses of variance (ANOVA) were conducted to investigate any differences among these groups. A stepwise regression model was conducted with GPCK as the dependent variable, and the GTBS constructs and the remaining TPACK-G constructs and demographic factors as independent variables.
Results

What are Taiwanese in-service teachers’ GTBS and TPACK-G perceptions in relation to their teaching level?

A series of independent *t*-tests was conducted to compare different teachers’ teaching beliefs about GTBS. As shown in Table 1, significant differences were identified in the dimensions of Belief (*t* = 4.06, *p* < .001), Confidence (*t* = 4.97, *p* < .001), and Motivation (*t* = 3.59, *p* < .001). The elementary school teachers outperformed the middle school teachers on Belief (Mean = 4.65; SD = 1.09), Confidence (Mean = 4.62; SD = 1.25), and Motivation (Mean = 3.95; SD = 1.23). That is, in comparison with the middle school teachers, the elementary school teachers were more likely to utilize games in their teaching. They believed that games could offer more learning opportunities and increase the effectiveness of their teaching. Table 1 also shows the results of the independent *t*-test analysis of the teachers’ TPACK-G. As shown, two statistical differences were found in GPK (*t* = 4.48, *p* < .001) and GPCK (*t* = 4.22, *p* < .001). The elementary school teachers outperformed the middle school teachers in terms of GPK (Mean = 4.63; SD = 1.32) and GPCK (Mean = 4.55; SD = 1.36). This suggests that the elementary school teachers were more inclined to enhance their students’ learning by using digital games. In addition, they were more likely to combine teaching subject, digital games and teaching approaches appropriately. However, no statistically significant difference was found in GK and GCK.

What are Taiwanese in-service teachers’ GTBS and TPACK-G perceptions in relation to their gender?

Teachers’ GTBS and TPACK-G were further investigated by gender. No statistically significant difference was identified in the teachers’ GTBS. The only statistically significant difference in TPACK-G was the GK factor (*t* = 4.15, *p* < .001). Male teachers’ GK score (Mean = 4.60, SD = 1.41) was higher than that of female teachers (Mean = 3.87, SD = 1.37), which suggests that male teachers tend to learn digital games more easily and have the technical skills to play digital games more effectively. That is, no significant difference was found in male and female teachers’ teaching beliefs about GBL teaching approaches, although they had significantly different self-efficacies regarding game playing knowledge.

What are Taiwanese in-service teachers’ GTBS and TPACK-G perceptions in relation to their age?

A series of ANOVA analyses were conducted to examine any statistically significant difference in GTBS and TPACK-G of the participants with different age groups. The results show that no statistically significant difference was found in the participants’ GTBS. For TPACK-G (See Table 2), two statistically significant differences identified were GK (*F* = 11.97, *p* < .001) and GCK (*F* = 3.18, *p* < .05). For GK, the results of the post-hoc tests indicate that teachers in the age group of 21 to 30 (Mean = 4.83, SD = 1.11) outperformed the other two groups (Mean = 4.10, SD = 1.48, for group 31-40; Mean = 3.72, SD = 1.37 for group over 40). This implies that in comparison with older teachers, the younger teachers tended to have higher self-efficacy for using games to represent the teaching content and learning digital games. Regarding GCK, the results of the post-hoc tests show that those aged 21 to 30 (Mean = 4.98, SD = 1.03) outperformed those over 40 (Mean = 4.44, SD = 1.09).
1.37). This finding suggests that when it comes to using games to represent subject knowledge, younger teachers were more confident than those who were over 40.

Table 2. ANOVA analysis of teachers’ GTBS and TPACK-G by different age levels

| TPACK-G | Age level (Mean, SD) |  |  |  |  |  |
|---------|----------------------|-------------------------------------------------|------------------|-----------------|------------------|
|         | (1) 21-30 (N = 50)   | (2) 31-40 (N = 119) | (3) over 40 (N = 139) |  |  |  |
| Belief  | 4.40 (.97)           | 4.34 (.95)           | 4.54 (1.17)           | 1.16 |  |  |
| Confidence | 4.48 (1.24)       | 4.22 (1.24)           | 4.31 (1.31)           | .72  |  |  |
| Motivation | 3.94 (1.17)       | 3.60 (1.14)           | 3.78 (1.24)           | 1.60 |  |  |
| GK      | 4.83 (1.11)           | 4.10 (1.48)           | 3.72 (1.37)           | 11.97*** | (1)>(2)** | (1)>(3)** |
| GCK     | 4.98 (1.03)           | 4.64 (1.32)           | 4.44 (1.37)           | 3.18*  | (1)>(3)* |  |
| GPK     | 4.47 (1.41)           | 4.28 (1.37)           | 4.31 (1.42)           | .31   |  |  |
| GPCK    | 4.57 (1.22)           | 4.16 (1.44)           | 4.20 (1.48)           | 1.54  |  |  |

Note. *p < .05, **p < .01, ***p < .001.

What are Taiwanese in-service teachers’ GTBS and TPACK-G perceptions in relation to their teaching experience?

This study divided the participants’ teaching experience into three levels: 0-10 years (n = 124), 11-20 years (n = 105) and over 20 years (n = 84). A series of ANOVA analyses was conducted to examine teachers’ teaching beliefs according to these three levels. As shown in Table 3, two statistically significant differences were found for Belief (F = 4.07, p < .05) and Motivation (F = 4.82, p < .01). For Belief, the results of the post hoc tests further showed that teachers with teaching experience of over 20 years (Mean = 4.62, SD = 1.10) significantly outperformed those with teaching experience ranging from 11 to 20 years (Mean = 4.21, SD = 1.12). In terms of Motivation, the results display that teachers whose teaching experience was less than 10 years (Mean = 3.92, SD = 1.13) scored higher than those whose teaching experience ranged from 11 to 20 years (Mean = 3.45, SD = 1.28).

As shown in Table 3, years of teaching experience played a role in GK (F = 19.16, p < .001), GCK (F = 6.51, p < .01), GPK (F = 3.18, p < .05) and GPCK (F = 4.41, p < .05). With a series of Scheffé tests, it was further indicated that teachers with teaching experience of less than 10 years tended to have higher self-efficacy in their GK, GCK, GPK and GPCK than those with teaching experience of more than 10 years. This implies that younger teachers tended to have higher confidence in appropriately integrating content, games and pedagogy for promoting students’ learning.

Table 3. ANOVA analysis of teachers’ GTBS and TPACK-G by years of teaching experience

<table>
<thead>
<tr>
<th>GAS</th>
<th>(1) 0-10 (N = 124)</th>
<th>(2) 11-20 (N = 105)</th>
<th>(3) over 20 (N = 84)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief</td>
<td>4.54 (98)</td>
<td>4.21 (1.12)</td>
<td>4.62 (1.10)</td>
<td>4.07** (3)&gt;(2)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>4.33 (1.26)</td>
<td>4.32 (1.31)</td>
<td>4.29 (1.25)</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>3.92 (1.13)</td>
<td>3.45 (1.28)</td>
<td>3.84 (1.17)</td>
<td>4.82** (1)&gt;(2)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GK</td>
<td>4.62 (1.21)</td>
<td>3.76 (1.45)</td>
<td>3.57 (1.38)</td>
<td>19.16*** (1)&gt;(2)<strong>; (1)&gt;(3)</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCK</td>
<td>4.92 (1.13)</td>
<td>4.37 (1.48)</td>
<td>4.40 (1.22)</td>
<td>6.51** (1)&gt;(2)<strong>; (1)&gt;(3)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPK</td>
<td>4.51 (1.31)</td>
<td>4.06 (1.49)</td>
<td>4.40 (1.35)</td>
<td>3.18* (1)&gt;(2)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPCK</td>
<td>4.52 (1.28)</td>
<td>3.97 (1.52)</td>
<td>4.21 (1.47)</td>
<td>4.41* (1)&gt;(2)**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001.

How do demographic factors and GTBS (Belief, Confidence, Motivation) and TPACK-G constructs (GK, GPK, and GCK) predict in-service teachers’ GPCK?

Table 4 shows that age and teaching experience were significantly highly related (r = .90, p < .001). Thus, both factors were also negatively related to the GK and GCK constructs (r = -.014 ~ -.32, p < .05). That is, the older
the teachers and the more teaching experience they had, the less their self-efficacy for GK and GCK. No significant relation was identified among age, teaching experience and the GTBS constructs. In addition, both GTBS and TPACK-G were significantly positively related. Among these, Confidence, Motivation, GK, GCK and GPK had stronger positive relations with GPCK, all above 0.60, suggesting large effect size coefficients (Cohen, 1988).

Table 4. Correlation among GTBS, TPACK-G, and demographics variables

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Teaching experience</th>
<th>Belief</th>
<th>Confidence</th>
<th>Motivation</th>
<th>GK</th>
<th>GCK</th>
<th>GPK</th>
<th>GPCK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>.90***</td>
<td>.07</td>
<td>.00</td>
<td>-.03</td>
<td>-.31***</td>
<td>-.14*</td>
<td>-.03</td>
<td>-.09</td>
<td></td>
</tr>
<tr>
<td><strong>Teaching experience</strong></td>
<td>1</td>
<td>.05</td>
<td>-.03</td>
<td>-.06</td>
<td>-.32***</td>
<td>-.17**</td>
<td>-.05</td>
<td>-.10</td>
<td></td>
</tr>
<tr>
<td><strong>Belief</strong></td>
<td>1</td>
<td>.29***</td>
<td>.67***</td>
<td>.28***</td>
<td>.36***</td>
<td>.46***</td>
<td>.48***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Confidence</strong></td>
<td>1</td>
<td>.47***</td>
<td>.44***</td>
<td>.46***</td>
<td>.61***</td>
<td>.65***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Motivation</strong></td>
<td>1</td>
<td>.39***</td>
<td>.40***</td>
<td>.62***</td>
<td>.64***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GK</strong></td>
<td>1</td>
<td>.65***</td>
<td>.58***</td>
<td>.60***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GCK</strong></td>
<td>1</td>
<td>.66***</td>
<td>.62***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GPK</strong></td>
<td>1</td>
<td>.86***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GPCK</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001.

Stepwise regression analyses were conducted to identify the predictive effects of the GTBS, TPACK-G, and demographic factors (teaching level, gender, age and teaching experience) on GPCK. That is, GPCK served as the dependent variable and the remaining TPACK-G constructs, GTBS, and demographic factors were processed as the predictor variables. As shown in Table 5, stepwise regression of the models was statistically significant with an adjusted R² of 0.78. Among the predictors, GPK was the critical factor as it explained 73% of the total variance in Model 1. The addition of GK, GCK, Motivation, and Confidence merely increased 5% (see Model 4). According to the Beta values, all the variables had positive prediction of GPCK. The regression models also indicated that the demographic variables of teaching level, gender, age, and teaching experience were not significant predictors of GPCK.

Table 5. Stepwise regression models for GPCK

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>Std. error</th>
<th>Beta</th>
<th>t</th>
<th>F</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(.Constant)</td>
<td>.46</td>
<td>.14</td>
<td>.86</td>
<td>3.35***</td>
<td>826.95</td>
</tr>
<tr>
<td>GPK</td>
<td>.87</td>
<td>.03</td>
<td>.73</td>
<td>28.76***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(.Constant)</td>
<td>.02</td>
<td>.15</td>
<td>.73</td>
<td>0.11</td>
<td>475.29</td>
</tr>
<tr>
<td>GPK</td>
<td>.74</td>
<td>.04</td>
<td>.21</td>
<td>20.36***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>.24</td>
<td>.04</td>
<td>.18</td>
<td>5.83***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(.Constant)</td>
<td>-.17</td>
<td>.16</td>
<td>.04</td>
<td>-1.06</td>
<td>335.31</td>
</tr>
<tr>
<td>GPK</td>
<td>.67</td>
<td>.04</td>
<td>.14</td>
<td>16.52***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>.21</td>
<td>.04</td>
<td>.18</td>
<td>5.21***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>.16</td>
<td>.04</td>
<td>.14</td>
<td>3.76***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(.Constant)</td>
<td>-.31</td>
<td>.16</td>
<td>.60</td>
<td>-1.94</td>
<td>265.10</td>
</tr>
<tr>
<td>GPK</td>
<td>.61</td>
<td>.04</td>
<td>.60</td>
<td>14.23***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>.19</td>
<td>.04</td>
<td>.17</td>
<td>4.76***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>.16</td>
<td>.04</td>
<td>.13</td>
<td>3.75***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GK</td>
<td>.12</td>
<td>.03</td>
<td>.12</td>
<td>3.65***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001.

Discussion

Game-based learning, a specific pedagogical environment, has been receiving growing attention in the last decade. Many researchers (Eastwood & Sadler, 2013; Kenny & McDaniel, 2011) have indicated that teachers play a dominant role in game-based learning. Thus, probing practitioners’ beliefs and perceptions of teaching with games has become essential. This study investigated Taiwanese in-service teachers’ teaching beliefs about game-based learning (GTBS) and their TPACK-G perceptions. The results of the stepwise regression models suggest that the teachers’ GPK plays an essential role in explaining their GPCK, accounting for 73% of the variance explained. This finding is resonant with prior research (Hsu et al., 2013) investigating 352 in-service preschool teachers’ technological pedagogical content knowledge of games. Similarly, the results also identified
a positive relationship between GPK and GPCK. Prior research has highlighted the importance of embedding pedagogy into game-based learning. According to Kiili (2007), playing educational games is likely to support the practice of learning factual information since the players might merely exhibit trial-and-error behaviors. Thus, adding pedagogy into game-based learning may enhance players’ engagement so as to achieve meaningful learning (Hsu, Tsai, & Wang, 2012), such as self-explanation (Adams, & Clark, 2014; Johnson & Mayer, 2010; Hsu et al., 2012), prediction-observation-explanation (POE) (Hsu, Tsai, & Liang, 2011), and problem-solving (Chang, Wu, Weng, & Sung, 2012; Kiili, 2007). Additionally, a recent study (Evans, Nino, Deater-Deckard, & Chang, 2015) used the TPACK framework to evaluate a learning game implementation. They found that while implementing games in class, teachers were able to have their students work in pairs and apply the concepts just learned in the gaming scenarios. Furthermore, traditional tests or quizzes were replaced with more innovative assessments, that is, playing games. Teachers could collect feedback or identify students with learning difficulties through the log files. As GPK is a fundamental element in developing GPCK, any strategies that enhance linkages between games and pedagogy, such as the aforementioned approaches, need to be pinpointed clearly for teachers.

In addition to GPK, the findings of this study show that motivation, confidence and GK are the significant variables that predict GPCK. Although many researchers (Celik & Yesilyurt, 2013; Sang et al., 2011; Teo, 2009) have found that teachers’ confidence and motivation would impact their ICT usage in class, rarely have the current TPACK studies investigated how teachers’ confidence and motivation relate to their perceptions of TPACK, particularly games (e.g., GPCK). The findings of this study can bridge this gap. Moreover, direct influence of in-service teachers’ TK on their TPACK has been identified in many TPACK studies (Graham et al., 2009; Koh et al., 2013; Koh, Chai, & Tsai, 2014). This study extends the findings of the aforementioned research by establishing a similar link to teachers’ classroom use of games. The present study also found that teachers’ perceptions of GPCK did not appear to be predicted by the demographic variables (e.g., teaching level, teaching experience, age and gender). This finding is resonant with the results of the prior research, suggesting that in-service teachers’ age and gender (Koh et al., 2014) or age and teaching experience (Chai et al., 2011) have no significant influence on their TPACK. In contrast, some studies still indicated that gender differences exist in in-service teachers’ TPACK (Lin et al., 2013). Teaching level and teaching experience play significant roles in predicting in-service teachers’ TPACK (Koh et al., 2014). The older the in-service teachers are, the less confidence they have in their TPACK (Lee & Tsai, 2010). These contradictory findings warrant further investigation in future studies.

This study found that the elementary school teachers tended to have more favorable Belief, Confidence, Motivation, GPK and GPCK than the middle school teachers (see Table 1). A selective interview was conducted to probe teachers’ beliefs about game-based learning. Most of the middle school teacher interviewees replied that due to the tight teaching schedule, they felt too rushed to find any spare time for integrating game-based learning into their classes. This, in comparison with elementary school teachers, might lead to their low confidence in GPK and GPCK together with their confidence and motivation for using game-based learning. Conversely, the elementary school teachers, even the more senior teachers, tended to show great interest in and willingness to utilize games to facilitate students’ learning. This finding is in contrast with Koh et al.’s (2014) study which investigated in-service teachers’ TPACK in Singapore. They found that due to the differences in college majors, secondary school teachers who were subject specialists and used ICT in a specific area tended to have higher confidence in their TPACK. However, the primary school teachers who were subject specialists and approached TPACK across various subjects were likely to have lower confidence in their TPACK. These inconsistent findings could be due to the cultural differences between the school teachers in Taiwan and Singapore. Another possible explanation could be that higher TPACK accompanies higher teaching loads, especially for teaching in a gaming context. Future studies can further explore the social, cultural and systematic issues in teachers’ teaching beliefs, motivation and self-efficacy for using games for instruction.

Finally, the results showed that a gender difference only occurred in GK. That is, the male teachers tended to have more confidence in their GK than the females. This corresponds with previous studies (Jang & Tsai, 2013; Lin et al., 2013) which revealed that in-service male teachers perceived more confidence in their TK. Regarding teaching experience, previous studies have shown that more experienced teachers were inclined to demonstrate higher TPACK than novice teachers (Jang & Tsai, 2012). Conversely, this study found that teachers with teaching experience of less than 10 years tended to perceive higher self-efficacy in their GK, GCK, GPK and GPCK than those with teaching experience of 11 to 20 years (see Table 1). Indeed, prior research has indicated that preservice teachers’ lack of teaching experience not only leads to their incapability of integrating technology into teaching (Ozgun-Koca, Meagher, & Edwards, 2010) but also the development of a new knowledge base (e.g., TPK) (Pamuk, 2012). Once they have accumulated some teaching experience, these younger teachers could be expected to confidently demonstrate a certain level of knowledge of game-based teaching.
In conclusion, this study found that the in-service teachers’ GPCK plays the most dominant role in predicting their GPCK. Other significant variables include motivation, confidence and GK, instead of teaching level, teaching experience, age and gender. In comparison with the middle school teachers, the elementary school teachers tended to have more favorable Belief, Confidence, Motivation, GK and GPCK. Younger teachers whose teaching experience was less than 10 years tended to perceive higher self-efficacy in their GK, GK, GPCK and GPCK than those whose teaching experience ranged from 11 to 20 years.

Acknowledgments

Funding of this research work was supported by the Ministry of Science and Technology, Taiwan, under grant number MOST 103-2511-S-011-002-MY3, MOST103-2511-S-011-005-MY3 & MOST103-2511-S-020-001-MY3.

References


Erdogan, A., & Sahin, I. (2010). Relationship between math teacher candidates’ technological pedagogical and content knowledge (TPACK) and achievement levels. Procedia - Social and Behavioral Sciences, 2, 2707-2711. doi:10.1016/j.sbspro.2010.03.400


**Appendix 1**

Belief 1: Digital games can improve teaching effectiveness.
Belief 2: Using digital games makes teaching easy.
Belief 3: Digital games can provide students with problem solving opportunities.

Confidence 1: It is difficult for me to integrate digital games into my instruction.
Confidence 2: I have no idea how to integrate digital games into my curriculum.
Confidence 3: I always feel frustrated when I use digital games in my classroom.

Motivation 1: I will not consider using digital games in my courses.
Motivation 2: I plan to use digital games in my classroom.
Motivation 3: When I prepare my teaching plans, I will link my curriculum with digital games.

*Note.* Confidence 1~3 and Motivation 1 were scored in reverse.

**Appendix 2**

GK1: I can learn digital games easily.
GK2: I can familiarize myself with the game interface.
GK3: I have the technical skills to play digital games effectively.
GK4: I know how to search for and download digital games.

GCK1: I can identify the knowledge related to the subject matter in the digital games.
GCK2: I can tell whether the digital games represent the targeted subject matter knowledge.
GCK3: I can identify whether the core concepts of the subject matter knowledge are displayed in the digital games.
GCK4: I can identify whether the subject matter knowledge is applied in digital games.

GPK1: I know how to use the characteristics of digital games to support teaching.
GPK2: I know the relevant instructional strategies of digital games.
GPK3: I know how to integrate digital games into teaching.

GPCK1: I can teach lessons that appropriately combine my teaching subject, digital games and teaching approaches.
GPCK2: I can craft real world problems about the content knowledge and represent them through digital games to engage my students.
GPCK3: I can select digital games to use in my classroom that enhance what I teach, how I teach and what students learn.
Wikis for a Collaborative Problem-Solving (CPS) Module for Secondary School Science

Dorothy DeWitt1, Norlidah Alias1*, Saedah Siraj1 and Jonathan Michael Spector2

1Department of Curriculum & Instructional Technology, Faculty of Education, University of Malaya, Kuala Lumpur, Malaysia // 2Department of Learning Technologies, College of Information, University of North Texas, Denton, Texas, USA // dorothy@um.edu.my // drnorlidah@um.edu.my // saedah@um.edu.my // mike.spector@unt.edu

*Corresponding author

(Submitted May 19, 2015; Revised October 26, 2015; Accepted May 20, 2016)

ABSTRACT

Collaborative problem solving (CPS) can support online learning by enabling interactions for social and cognitive processes. Teachers may not have sufficient knowledge to support such interactions, so support needs to be designed into learning modules for this purpose. This study investigates to what extent an online module for teaching nutrition in secondary school science, using a wiki for CPS, enables interactions, and social and cognitive processes. The module was implemented with 31 volunteer participants. Data collected from the online communications was analyzed for the types of interactions and processes based on the Community of Inquiry Framework. This was triangulated using transcripts of interviews with students. In addition, pretests and posttests were conducted to determine whether the learning outcomes were achieved. Analysis of the online communications showed that the interactions were mainly between learner and content (64.4%), with a large portion of cognitive processes (69.3%) but little social (4.0%), attitudes (9.9%), teaching processes (12.9%) and noise (5.9%). The findings suggest that the module could be used to improve outcomes of learning and encourage interactions for cognitive processes and online presences. The findings may provide insights in encouraging CPS for learning science online.

Keywords

Collaborative problem-solving, Community of Inquiry Framework, Online pedagogy, Wiki

Introduction

Collaborative problem-solving (CPS) in online environments can afford high levels of interactions among learners which can serve to develop critical thinking skills (Kim & Song, 2006; Manathunga & Hernández-Leo, 2015). The authentic and non-linear approach in problem-solving with sufficient support from peers, instructors and the learning environment can support learning (Cela, Sicilia & Sánchez, 2015; Heo, Lim, & Kim, 2010; Puntambekar, 2006; Siraj & Norman, 2012; Vaughan & Garrison, 2005). Collaborative learning and problem-solving have also been shown to be effective in science education, yet few studies have examined online collaboration and problem-solving in science (Slotta & Linn, 2000; van der Spa, 2004; Turcotte, 2012). In this paper, the integration of collaborative problem solving in a wiki to support science education is addressed. What is perhaps unique is the investigation of collaboration and interaction activities in the wiki and the measurements made to determine how CPS activities in the wiki may have contributed to learning science.

In Malaysia, mathematics and science achievement have declined in the last 5 years (Ministry of Education Malaysia, 2013). In order to assist in improving the education system, several shifts have been identified in the Malaysia Education Blueprint 2013-2025 to ensure that students could compete globally (MOE, 2013). One of the shifts identified is for the use of Information and Communication Technology (ICT) using new pedagogies to deliver quality education. However, in Malaysian schools, CPS and new pedagogies are rarely used in teaching science in secondary schools (Mohamad Said, 2011). Teachers generally teach facts and concepts, and emphasize memorization and recall (DeWitt & Siraj, 2008; Phang, Abu, Ali, & Salleh, 2012). Teachers perceive there is a lack of time for completing the syllabus in the classroom and rarely use ICT for instruction (DeWitt & Siraj, 2008). In addition, to be able to employ ICT for instruction, teachers require knowledge of online pedagogies. The principles of an online pedagogy include: (a) students are responsible for their learning as they search for materials and build their knowledge, and (b) students should collaborate and interact while working on projects (Pelz, 2004).

This study investigates to what extent CPS can support those principles of online pedagogy in the context of teaching science in a Malaysian school. This study is a single case with a specific cultural context; however, it can assist in the implementation of online CPS in other schools in Malaysia and in the region. A module using a
wiki was used in this study. Due to the lack of time, the CPS module was implemented outside the classroom after school hours. A wiki was chosen to minimize the burden on teachers while emphasizing collaborative problem solving. It is assumed that successful implementation requires knowledgeable and engaged teachers.

In this study, collaborative learning is defined as new knowledge, skills and attitudes acquired as a result of interactions in a group (Jonassen, Lee, Yang, & Laffey, 2005). Collaborative mLearning occurs when new knowledge and skills can be accessed anywhere and anytime in online environments, and the acquisition of knowledge is a result of online interactions in a group (DeWitt & Siraj, 2008). These online interactions include postings on wikis, discussion forums and chat groups.

**Purpose of the research**

This study investigates to what extent a module using a wiki for online CPS is useful for learning science. The research questions were:
- What kinds of student interactions occur in a wiki for learning science?
- To what extent does CPS in a wiki encourage social interactions, and cognitive processes?
- To what extent does the module improve achievement?

**Interactions for understanding in science**

In science instruction, providing opportunities for interaction and collaboration is important for learners to acquire scientific knowledge (Etkina, Mestre, & O’Donnell, 2005; Hogan & Fisher Keller, 2005; Osbourne & Henessy, 2003). Scientific verbal knowledge assists the construction of meaningful language to discuss and give feedback during communication (Hogan & Fisher Keller, 2005).

In collaborative learning, learners communicate and resolve differences during discussions to reach mutual understanding (Jonassen et al., 2005; Palloff & Pratt, 1999). This process of collaboration and resolving differences can develop learners’ critical thinking skills (Kim & Song, 2006). The learners’ interactions, from learner to learner; learner with the instructor; and learner with the learning materials, can enhance understanding of science concepts and principles (Kim & Song, 2006). However, interactions alone do not ensure effective collaboration among the groups; tasks and guidance are needed to ensure effective online learning collaboration (Collazos, Guerrero, Pino, & Ochoa, 2003).

**Collaborative problem-solving (CPS)**

Collaborative learning is a natural learning process (Vygotsky, 1962; 1988); it occurs as a result of learners’ responses and interactions within the community of learners (Barrows & Tamblyn, 1980; Johnson & Johnson, 2004; Seel, 2003; Spector, 2004). Learning is still an individual process, but is influenced by group and interpersonal interactions (Kaye, 1992; Manathunga & Hernández-Leo, 2015). Each individual brings prior knowledge and perspectives to the learning environment, and collaborative learning allows these different perspectives to form a shared knowledge within the community as a result of social interactions (Jonassen et al., 2005; Kaye, 1992; Palloff & Pratt, 1999).

The dialogues and interactions with other learners and the instructor can have a positive influence on learning. When collaborative activities effectively promote learning, positive interdependence develops as group members work together on shared tasks and resolve conflicts while working toward a common goal (Collazos et al., 2003). However, negative interdependence may occur when individuals have little interaction and do not focus on the goals of the group.

Problem-solving tasks allow learners to take responsibility for their own learning, which is the first online pedagogical principle (Barrows & Tamblyn, 1980; Johnson & Johnson, 2004; Pelz, 2004; Seel, 2003; Spector, 2004). In line with this principle, learning should be learner-directed and the instructors’ role is to provide the learning tasks and support (Hannafin et al., 2009; Johnson & Johnson, 2004; Jonassen et al., 2005; Pelz, 2004). Learners should be allowed the responsibility to determine their own learning (Hannafin et al., 2009).
Ill-structured problem-solving tasks are meaningful and authentic. Creative ideas are generated and higher order thinking increased during the interactions in problem-solving activities (Hannafin et al., 2009; Heo et al., 2010; Jonassen et al., 2005; Palloff & Pratt, 1999; Puntambekar, 2006; Siraj & Norman, 2012; Vaughan, 2010; Woo & Reeves, 2007).

Lack of adequate prior knowledge may contribute to learners’ misconceptions (Hannafin et al., 2009; Kirschner, Sweller, & Clark, 2006). Hence, the importance of the second online pedagogical principle, interaction, for scaffolding among peers, with an instructor, or with the content and technology, through system prompts, support the learner in solving ill-structured problems (Hannafin et al., 2009; Pelz, 2004; Vaughan & Garrison, 2005). Scaffolding, through interactions, could assist learners in constructing accurate representations of solution to problems (Belland, 2010).

CPS promotes strategies for positive interdependence. Scaffolding needs to be incorporated with the suitable problem tasks as learners help each other in the collaborative classrooms, aided by their instructor to improve learning achievement (Collazos et al., 2003; Whipp & Lorentz, 2009). The instructor should not intervene too much in the online interactions and instead allow learners to be independent and learn from their peers; however, the groups still need to be monitored and some guidance and scaffolding should be provided to help students reason through ill-structured problem tasks (An, 2010; Collazos et al., 2003).

In this study, the problem-solving task required learners to be responsible for their learning as they search for information and post solutions on the wiki. Support was provided in the content and reference learning materials and tools; support came from peers and the instructor.

Wikis for collaboration

Wikis and discussion forums have been used for collaborative learning in an inquiry-based teaching approach in science (Slotta & Linn, 2000; van der Spa, 2004; Turcotte, 2012). Wikis enable high levels of collaboration and stronger mutual engagement in the community of learners who interact with each other (Lu, Lai & Law, 2010). When used for problem-solving and social interaction, wikis can encourage reflection, critical thinking, motivation and understanding (Altanopoulou, Tselios, Katsanos, Georgoutsou, & Panagiotaki, 2015; Biasutti & El-Deghaidy, 2012; Driscoll, 2007; Higdon & Topaz, 2009; Jonassen et al., 2005; Kaye, 1992; Lu, Lai & Law, 2010; Osbourne & Hennessy, 2003; Siraj & Norman, 2012).

Wikis have been used for documenting scientific inquiry in secondary school science (DeWitt, 2010; DeWitt, Alias & Siraj, 2015; Hannafin et al., 2009). Designed for collaboration, this asynchronous communication tool can transform knowledge as learners generate, share and reshape knowledge (Bonk, Lee, Kim, & Lin, 2009). Learners participate in active learning and contribute to building knowledge on the wiki (Bonk et al., 2009; Pifarré & Li, 2012). When information is obtained and evaluated, the learner assimilates and applies the knowledge while solving problems (Biasutti & El-Deghaidy, 2012). Simultaneously, the learner’s peers and instructors scaffold him or her to achieve expert status (Pifarré & Li, 2012; Whipp & Lorentz, 2009).

Wikis can enable development of reading and writing skills (Imperatore, 2009; Shihab, 2009). When CPS is employed in wikis, it can assist in developing project management skills as learners are responsible for managing their teams, assigning roles and responsibilities, improving motivation, and monitoring group efforts (Biasutti & El-Deghaidy, 2012; Ertmer et al., 2011; Hannafin et al., 2009).

Some studies have shown low interaction between learners involved in activities on wikis; however, on investigation the lack of interaction was due to the lack of knowledge (Ertmer et al., 2011; Huang, 2010). Students feared negative judgment; the decreased confidence and expectancy for success was the problem (Ertmer et al., 2011). As a result, it is important to ensure that students have the requisite knowledge need to make positive contributions in the wiki.

The full potential of wikis in learning science needs to be investigated further to determine its usefulness in supporting an online pedagogy which enables self-directed learning, promotes interactions, and maintains an online presence with social and cognitive processes for learning (Ertmer et al., 2011). As the type of interactions in a wiki used for CPS may differ from other Web-based learning tools, there is a need to investigate the processes present in a wiki (Huang, 2010; Pelz, 2004). This study represents one step in that direction.

146
Processes in online interactions

A progressive pedagogy encourages higher order thinking as instructors use arguments, debates, dilemmas for conceptual conflicts, sharing of ideas, reflection and problem-solving in instruction and encourage positive interdependence in collaborative groups (Collazos et al., 2003; Kanuka, 2010; Spector, 2015). Pelz (2004) summarized three principles for effective online pedagogy: students are responsible for their learning as they independently lead discussions, access materials, assist their peers and perform self-assessment; interactivity is maintained through collaborative activities; and online presences is enabled through social, cognitive, and teaching presences. In this study, the collaborative task allowed students to be self-directed learners, and an attempt to measure the interactions and presences were made.

However, collaborative processes are difficult to measure. Most researchers focused on the outcomes of the collaborative process using tests, while others some have attempted to measure group performance. The Index of Collaboration uses metrics such as number of errors during the activity, ability to solve problems, number of queries, explicit use of a strategy, maintaining and communicating the strategy and the different forms of interactions (Collazos, Guerrero, Pino, Renzi, Klobas, Ortega, Redondo, & Bravo, 2007). These metrics were used as indicators, examples, for developing strategies and intra-group cooperation, reviewing success criteria, and monitoring group performance (Collazos et al., 2007). Attitudes and social aspects were not measured. However, Collazos et al. (2007) did recommend that other factors be studied in future: subject of the collaborative task, age, culture, and types of groups.

In CPS, the type and quality of interactions could influence learning. Originally, three types of online interaction were classified: learner-content, learner-learner, and learner- instructor (Moore & Kearsley, 2005). Later, another category of online interaction where the learner interacts with the technology medium or learner-interface interaction, was included (Hillman, Willis, & Gunawardena, 1994; Thurmond & Wambach, 2004). Meaningful interactions are important as solving authentic and meaningful problems will enhance learning (Kim & Song, 2006; Woo & Reeves, 2007).

Table 1. Categories for analysis of online interactions

<table>
<thead>
<tr>
<th>Categories of interactions</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Process:</td>
<td></td>
</tr>
<tr>
<td>- Triggering</td>
<td>Critical thinking components: triggering, exploration, integration and resolution</td>
</tr>
<tr>
<td>- Exploration</td>
<td>Phrases that encourage thinking about or posing a problem</td>
</tr>
<tr>
<td>- Integration</td>
<td>Phrases that are related but are not supported, phases which indicate search for information</td>
</tr>
<tr>
<td>- Resolution</td>
<td>Phrases that build on previous messages and are supported. This is the development of a possible solution</td>
</tr>
<tr>
<td>Teaching Presence</td>
<td>Phrases that suggest a hypothesis and ways to test and assess the solution</td>
</tr>
<tr>
<td>Social Presence</td>
<td>Design of the learning experience, as well as delivery and facilitation of students (and teacher)</td>
</tr>
<tr>
<td>Attitude</td>
<td>Characteristics of social interaction such as cohesiveness.</td>
</tr>
<tr>
<td>Noise</td>
<td>Components that indicate the affective aspect of communication</td>
</tr>
<tr>
<td></td>
<td>Communications that cannot be categorized</td>
</tr>
</tbody>
</table>

Note. Adapted from Shedletsky (2010), and Pinzon-Salcedo et al. (2008).

Online interactions can be further analyzed to determine cognitive, social, and teaching processes, and attitudes (Moore & Kearsley, 2005). Some research reports show little evidence of the cognitive processes in students’ online communications (Garrison, Anderson, & Archer, 2010), but other research shows evidence of more critical thinking processes in online communications as compared to face-to-face communications (Shedletsky, 2010). These interactions can be analyzed using the Community of Inquiry (COI) Framework for social, cognitive and teaching presence (Garrison et al., 2010; Vaughan, 2010). Shedletsky (2010) included a fourth category, discourse. Further, Pinzon-Salcedo, Barros, Zarama, de Meza, Carulla, and Bejarano (2008) included two additional categories: (a) attitudes about interaction (affective aspect), and (b) noise (communications that could not be identified or placed in other categories). These two categories can be analyzed as part of the discourse (Shedletsky, 2010). Hence, the types of presence and processes can be investigated. In this study, social, cognitive and teaching presence, as well as attitudes and noise were analyzed in the wiki (see Table 1).

In this study, the interactions in the online environment were analyzed to determine the types of interactions and presences (Pelz, 2004) and the outcomes of learning were compared using pretests and posttests results.
Methodology

Significance of the study

The results of this study can be used to determine if CPS using wikis encourages an online pedagogy suitable for learning. Teachers could then use this approach in teaching online and developing positive interdependent tasks for self-directed learning, and enable interactions for an online presence for learning.

Scope and limitations

In this study, an urban secondary school with a multiracial population reflective of the multiracial communities in Malaysia was selected on the bases of representativeness and convenience. The sample of 31 student volunteers comprising of high, medium and low-achievers in science were selected to participate in the implementation of a science lesson on Food Classes. Lessons were designed and developed for evaluation in this study using a wiki for CPS. One limitation in this study is that only the online interactions in the wiki and in the discussion forum were captured and analyzed. There may have been face-to-face interactions, but these were not analyzed as it was not within the scope of the study. In addition, the focus of the study was on the interactions, presences, and achievement as an outcome of learning, and not on other metrics of collaboration.

Instruments

Firstly, the online interactions were analyzed and classified as learner-content, learner-learner, learner-instructor, and learner-interface. Next, the interactions were analyzed based on the categories of cognitive, social, and teaching presences, as well as attitude and noise (see Table 1). Coding of the online documentations on the wiki was done by the researcher based on these categories and reconfirmed by a co-researcher to determine if there were differences in coding. The differences were resolved through discussion.

The second instrument involved two sets of test questions consisting of simple open ended questions for pretest and posttest. This test was designed to determine students’ achievement in the Food Classes module. The 10 items in each instrument tested the same constructs, such as naming the foods in the different food classes. The instrument was content-validated by two Biology teachers with more than 10 years’ working experience, and is in use for testing understanding of the module’s content.

Implementation procedure

A CPS module using a wiki was developed for learning food classes. The participants could use the module either during, or after school hours, and laptops were made available for those who did not have access to computers. Students worked in groups of seven to eight on the wiki. The students who volunteered to use the module were given a face-to-face orientation to the module, which consists of a web page with links to learning resources and tools, and the problem tasks and a wiki was set up for each group for the task which was to analyze the food classes in a meal.

Different meals were assigned to the groups, and they were given three weeks to complete the task on their group wiki page. The group members could add pages, links and graphics to their wiki page. During the implementation, feedback on the task could be given on the wiki or through other forms of communication by other learners and the instructor. The instructor communicated online, but the students interacted both online and face-to-face in the school environment.

Results and discussion

Data collection and analysis

In this study, data were collected from communications on the wiki and discussion forums, journal records; interviews after task completion and tests (An, 2010; Whipp & Lorentz, 2009). The communications and solution displayed in the wiki were coded and analyzed according to the types of interactions (Hillman et al., 1994; Moore & Kearsley, 2005; Thurmond & Wambach, 2004) and further into processes (Pinzon-Salcedo et al.,...
In addition, randomly selected students were interviewed to probe further into the process and the problems encountered. The pretest and posttest scores were analyzed for significance using a paired t-test to compare means of their scores.

Types of interactions

The first research question to determine the types of interactions among students in the context of the study when using a wiki showed the interactions occurring in the wiki were mainly learner-content (see Table 2). There was little evidence of learner-learner interaction, and learner-interface interaction on the wiki.

<table>
<thead>
<tr>
<th>Types of interaction</th>
<th>f</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner–Content</td>
<td>29</td>
<td>(64.4)</td>
</tr>
<tr>
<td>Learner–Learner</td>
<td>3</td>
<td>(6.7)</td>
</tr>
<tr>
<td>Learner–Instructor</td>
<td>13</td>
<td>(28.9)</td>
</tr>
<tr>
<td>Learner–Interface</td>
<td>0</td>
<td>(0.0)</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>(100.0)</td>
</tr>
</tbody>
</table>

Note. f = frequency of interactions.

Lack of learner-learner interactions seemed to be because learners only posted information on the problem solution on the wiki. An example of interaction in one group discussed what they needed to do:

W: What do we need to do?
M: I think that they want us to classify the meals.
N: Should we classify the meals?

This lack of interaction on the wiki seems to be similar to other studies (Ertmer et al., 2011; Huang, 2010). Interviews suggested that the apparent lack of contribution in the wiki was because group discussions had already taken place in online chats and face-to-face discussions.

There was evidence of learner-instructor interaction through emails and text messages on Yahoo Messenger, an online chat application. The instructor had given suggestions through text-message chat on how to improve the task and assisted in solving a technical problem (Figure 1). Hence, the absence of interaction on the wiki did not mean there were no interactions.

There was little positive interdependence strategy for the instructor to redirect questions learners posed back to the individual, or to ask teams to seek help from other teams (Collazos et al., 2003). Learners seem to prefer interacting with the instructor. However, the groups seemed to achieve positive goal interdependence as there was motivation in working together for task completion, perhaps because there was reward independence due to the marks given for successful group work (Collazos et al., 2003).

As for learner-content interaction, participants collaborated on the tasks in most groups. However in all groups, there seemed to be one dominant member posting most of the solutions. This was because the groups had assigned one group member responsible for wiki posts (Ertmer et al., 2011),
Social and cognitive processes

To answer the second research question on social interactions and cognitive processes, analysis of the online communications transcripts was conducted. There were mostly cognitive processes (69.3%), with very little social (4.0%), attitude (9.9%), teaching (12.9%) and noise (5.9%) on the wiki (see Table 3). Analyzing the cognitive processes, hardly any triggering interactions to encourage thinking on the issues occurred on the wiki, but there was evidence of some exploration (3.0%) in searching and connecting information. A large proportion in all groups was integration (54.5%), in which messages are built on other messages in order to construct a solution, while resolution to assess the solution was much lower (11.9%).

Table 3. Categories of interactions on the wiki

<table>
<thead>
<tr>
<th>Categories of interactions</th>
<th>Group 1 (f (%)</th>
<th>Group 2 (f %)</th>
<th>Group 3 (f %)</th>
<th>Group 4 (f %)</th>
<th>Total (f %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Process:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Triggering</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>• Exploration</td>
<td>1 (33.3%)</td>
<td>0 (0%)</td>
<td>2 (66.7%)</td>
<td>0 (0%)</td>
<td>3 (3.0%)</td>
</tr>
<tr>
<td>• Integration</td>
<td>22 (40.0%)</td>
<td>6 (10.9%)</td>
<td>5 (9.1%)</td>
<td>22 (40.0%)</td>
<td>55 (54.5%)</td>
</tr>
<tr>
<td>• Resolution</td>
<td>3 (25.0%)</td>
<td>6 (50.0%)</td>
<td>3 (25.0%)</td>
<td>12 (11.9%)</td>
<td></td>
</tr>
<tr>
<td>Teaching Presence</td>
<td>2 (15.4%)</td>
<td>5 (38.5%)</td>
<td>4 (30.8%)</td>
<td>13 (12.9%)</td>
<td></td>
</tr>
<tr>
<td>Social Presence</td>
<td>0 (0%)</td>
<td>1 (25.0%)</td>
<td>3 (75.0%)</td>
<td>4 (4.0%)</td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>10 (100.0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>10 (9.9%)</td>
<td></td>
</tr>
<tr>
<td>Noise</td>
<td>5 (83.3%)</td>
<td>1 (16.7%)</td>
<td>0 (0%)</td>
<td>6 (5.9%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>43 (42.6%)</td>
<td>13 (12.9%)</td>
<td>13 (12.9%)</td>
<td>32 (31.7%)</td>
<td>101 (100.0%)</td>
</tr>
</tbody>
</table>

The lack of triggering interactions could be attributed to the fact that the wiki was considered as a space to publish their completed task. It is believed that the triggering interactions occurred during the face-to-face interactions and was not captured online. This was reinforced during the interview when members clarified group discussions was done on other platforms before posting online.

Evidence of exploration was seen as in two groups. In one case, a tool for computing the number of calories in the different meal ingredients was shared: “I utilized the tool to compare nutrition data.” In another example, a statement on comparison of the type of flour was made to gather information for group agreement: “The recipe included 100gm of whole meal flour, that’s nearest to whole-grain wheat flour.”

There was integration recorded in all groups, as messages are built on other messages in order to construct a solution with the highest number in Groups 1 and 4 (40.0% each). The statements written showed the solution directly, example: “That contains 339 calories. There is 2g of fat and 73g carbohydrates, with 12g is dietary fiber and 14g of protein.”

Resolution, a higher level process in which the solution is assessed, was observed in three groups. In one example the learner made an assessment that there were other minerals they had not learnt in school.

There seems to be a lot of Vitamin B (Thiamin, riboflavin, niacin, folate and vitamin B6 and B12 - are all vitamin Bs), a little of vitamin E and K. It is also rich in iron and phosphorus, and a bit of calcium. There is also a lot more minerals which we do not learn about in school.

In another example, the food classes in the chapatti meal were discussed, and the learner concluded that it comprised all the food classes:

So in chapatti alone there is carbohydrate, protein, fat, vitamins A, B, E and K, minerals calcium, iron, phosphorous. Fiber and water too. I think most of the food classes are here.

Social interactions were few, and detected in two groups, for example:

M: I know it’s a bit late but I found out a lot about our meal. Nasi lemak is served for breakfast and is widely eaten by all races. The rice is served with anchovies and peanuts.
N: I don’t like it, it always gets stuck in my throat.
Where it occurred, social interactions were a positive interdependence for the group. However, fewer social processes online did not mean there was less interactions among group members, as the social processes seemed to have occurred outside the wiki.

On the other hand, the teaching process was evidenced in all groups when the instructor was asked questions on discussion forums and online chats, and when learners accessed content for learning. There was a lot of noise and attitude among the learners. The birthday wishes in Figure 2 was considered social, and noise. Another example, “Some just eat nasi lemak to save money.” Attitude can be seen from an elaborate description of the meal: “As its name suggests, the rice is cooked by steaming in coconut cream with pandan leaves to give a seductive fragrance. Spices such as ginger and lemon grass may be added for fragrance.”

![Figure 2. Screen capture showing noise and attitudes in the wiki (Group 2)](image)

To answer the research questions, there were cognitive and social processes occurring. Specifically, there seemed to be few triggering processes probably because the search for information was not required in the task. There were also not many explorations of the solution probably because discussions to explore the suggested solutions were done elsewhere, face-to-face or online, and were not captured on the wiki.

Hence, more opportunities for sharing and debating ideas online should be provided, such as an online chat platform for group members. This might also promote social processes which are important in encouraging a positive attitude among group members towards online learning for positive interdependence. Hence, the pedagogy should include opportunity for learners to explore, search and share information with group members assigned different resources to share in the group (Collazos et al., 2003).

The highest cognitive process, resolution, did occur in the wiki as learners assessed the solutions but only in three groups. Learners might not be able to demonstrate it effectively. In the wiki, a simple form of assessing the solution was indicated. In future, an assessment of their work should be included as part of the CPS task. In conclusion, social and cognitive processes did occur, but might not have been captured on the online interactions in the wiki. These processes were important and contributed to science learning.

**Learning effectiveness**

The final research question is whether interactions improved outcomes for learning science. The effectiveness of the CPS module was analyzed using paired t-test for 31 participants to determine if any significant differences existed between the mean pre and post test scores. Table 6 shows the results of t-test comparison of pre and posttest achievement. The findings showed that the CPS module is effective for teaching science. There is a significant difference between pretest (mean = 62.37, SD = 19.23) and posttest (mean = 82.55, SD = 12.78) scores, \( t(31) = 7.230, p < .05 \) (Table 4). The t-test indicates that the mean score of the posttest is significantly higher than for the pretest.

These findings are verified from the large number of cognitive interactions in the wiki. The learners demonstrated the different cognitive processes of exploration, integration and resolution in the online interactions on the wiki in most groups. Social interactions are important for learning science as these develop
learners’ cognitive skills (Karpov & Haywood, 1998). In this study, there was not much online social interaction recorded. However, the face-to-face peer interactions had facilitated the knowledge-building processes.

### Table 4. t-test comparison of pre and posttest scores

<table>
<thead>
<tr>
<th></th>
<th>Pretest (n = 31)</th>
<th>Posttest (n = 31)</th>
<th>t-value</th>
<th>p</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>62.37</td>
<td>82.55</td>
<td>7.230</td>
<td>&lt; .05</td>
<td>0.00</td>
</tr>
<tr>
<td>SD</td>
<td>19.23</td>
<td>12.78</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. n = number of participants.*

### Implications and conclusions

The learners in this study were novice users of the wiki and unfamiliar with ill-structured problem-solving processes. They were accustomed to traditional teacher-centered instruction. However, this did not seem to be a problem as regardless of their lack of experience in online CPS, the learners managed to present acceptable solutions on the wiki. In addition, they managed to participate in the tasks, which afforded the principles of effective online pedagogy. The learners were responsible for their learning and were self-directed, actively interacting and collaborating, and striving for social, cognitive or teaching presence (Pelz, 2004).

There did not seem to be many interactions in the wiki but it was not because of lack of knowledge as cognitive processes were detected on the wiki (Ertmer et al., 2011; Huang, 2010). The interactions between learners were conducted outside the wiki as well. Hence, future studies should be conducted to investigate these interactions as well. Similar studies could be conducted in a more controlled environment to capture both online and offline interactions for learning. Another alternative is to conduct the study with group members who are separated by distance, and unable to conduct face-to-face meetings except through online platforms.

The instructors could play an important role in scaffolding learners for positive interdependence toward achieving higher level cognitive skills and preparing learners for learning environments with ill-structured problem-solving. Further investigation is required to determine whether additional support strategies for positive interdependence and social interactions between the instructor and among learner-learner interaction could increase cognitive skills in such complex, ill-structured problem-solving environments.

Social interactions are important for learning science (Hogan & Fishkeller, 2005). In addition, interactivity and creating social presence is important for online pedagogies. During CPS, social interactions contribute to the acquisition of knowledge and skills through cognitive processes such as concept-formation, resolving differences and critical thinking (Tzeng, Chiang, Yang, & Huang, 2012; Karpov & Haywood, 1998; Kim & Song, 2006). Hence, there is a need for investigating the strategies for encouraging social interactions. In addition, future studies could be done to determine whether academic performance could improve with the use of wikis for CPS in other subjects besides science.

This study supports prior research and indicates that CPS can be conducted in a wiki (Altanopoulou et al., 2015; Bonk et al., 2009; Manathunga & Hernández-Leo, 2015; Pifarré & Li, 2012). In addition, a CPS module using a wiki can be effective for teaching science and developing cognitive skills and processes among novice learners. The interactions and presences on the wiki can help transform learning and develop understanding in online environments.

### References


How Augmented Reality Enables Conceptual Understanding of Challenging Science Content

Susan Yoon¹*, Emma Anderson¹, Joyce Lin² and Karen Elinich³

¹Graduate School of Education University of Pennsylvania, Philadelphia PA, USA // ²Knowles Science Teaching Foundation, Moorestown, NJ, USA // ³The Franklin Institute, Philadelphia, PA, USA // yoonsa@gse.upenn.edu // emmaa@gse.upenn.edu // joyce.lin@kstf.org // kelinich@fi.edu

*Corresponding author

(Submitted November 19, 2015; Revised January 28, 2016; Accepted July 28, 2016)

ABSTRACT
Research on learning about science has revealed that students often hold robust misconceptions about a number of scientific ideas. Digital simulation and dynamic visualization tools have helped to ameliorate these learning challenges by providing scaffolding to understand various aspects of the phenomenon. In this study we hypothesize that students acquire a more accurate understanding of the Bernoulli’s principle, a challenging science concept, by interacting with an augmented reality (AR) device. We show that even given a short period for investigation in a science museum, students in the AR condition demonstrate significantly greater gains in knowledge over students in the non-AR condition. Through interview responses, we further show that the AR affords greater ability to visualize details and hidden information to help students learn the science.

Keywords
Augmented reality, Challenging science content, Bernoulli’s principle

Introduction

Research on learning about science has revealed that students often hold robust misconceptions or naïve conceptions about a number of scientific ideas (Chi, 2005; Bransford, Brown, & Cocking, 1999). For example, studies on student understanding of the Bernoulli’s principle, which is the subject of our exploration, have shown that students find learning the content challenging due to, among other things, the counterintuitive experiences of pressure-related events observed in the real world (Stepans, 2003).

Digital simulation and dynamic visualization tools have helped to ameliorate these learning challenges by providing scaffolding (Honey & Hilton, 2011; Kim & Hannafin, 2011) to understand various aspects of phenomenon that may contribute to misconceptions. Related to this, a recent focus in the learning sciences has investigated how augmented reality (AR) tools can support science learning (Dunleavy, Dede, & Mitchell, 2009; Dunleavy & Dede, 2014; Klopfer & Squire, 2008). At its simplest, augmented reality describes systems that integrate computer-generated virtual elements or information (known as “digital augmentations”) with the real world environment (Zhou et al., 2008). By superimposing virtual elements onto the real world environments, AR allows users to experience and perceive the newly incorporated information as part of their present world, thereby enhancing their perception of the real world (Kirkley & Kirkley, 2004; New Media Consortium, 2012).

Everyday examples of AR include Google Effects in Hangouts, AR games for Nintendo 3DS, and Webcam Greeting cards from Hallmark.

Over the last 4 years, our project, Augmented Reality for Interpretive and Experiential Learning (ARIEL), has investigated optimal uses of AR in science museums (e.g., Yoon, Elinich, Wang, Steinmeier, & Tucker, 2012a; Yoon, Elinich, Wang, Van Schooneveld, & Anderson, 2013; Yoon & Wang, 2014), where misconceptions about science are rarely addressed. In this study, we hypothesized that the use of AR, because it provides a visualization of the underlying causal mechanisms, can assist students in developing a more accurate conception of Bernoulli’s principle. We found that after participating in brief, informal investigations of the principle at a science museum, students who interacted with an exhibit using AR were better able to understand the science than students in a non-AR condition. Findings from our interviews and surveys suggest that the tool supported students’ learning by revealing typically invisible features of the phenomenon.
Theoretical considerations

Common misconceptions and challenges associated with understanding Bernoulli’s principle

Bernoulli’s principle states that “when an incompressible, smoothly flowing fluid gains speed, internal pressure in the fluid decreases, and vice versa” (Hewitt, 2004). In other words, there is an inversely proportional relationship between fluid speed and pressure. When the fluid’s speed increases, the pressure drops. As it turns out, this is a conceptually challenging and counterintuitive idea to understand for students, who typically believe that when speed increases, so does the pressure (Faulkner & Ytreberg, 2011). Stepsen (2003) explains,

*Children learn from experience that when they blow on something – like a bubble or dandelion plume – it goes away. These experiences make it difficult to make sense of the fact that when you blow on a surface, it comes toward you, or that when you blow between things, they come together. These experiences make it difficult to accept the concept of Bernoulli’s Principle.* (p. 46)

In a test given to private and public 6th, 7th, and 8th grade Turkish students on the outcomes of discrepant events related to Bernoulli’s principle, Bulunuz, Jarrett, and Bulunuz (2009) found that the majority of students held incorrect conceptions of the phenomenon. Researchers gave a similar test to a group of pre-service elementary teachers, and results indicated that less than 50% of the teachers gave correct responses (Bulunuz & Jarrett, 2009). Even physics teachers claim to have unclear understandings of Bernoulli’s principle and sometimes avoid teaching the concept altogether to their students (Hewitt, 2004).

In examining misconceptions of pressure-related concepts, Basca and Grotzer (2001) organize the conceptual challenges into four categories, two of which have important implications for our research. The first difficulty that children have is that they tend to reason using obvious, rather than nonobvious, variables when determining the causes of pressure-related events. For example, de Berg (1995) found that when a syringe was compressed, 17- and 18-year-old students felt, and accurately identified, that the pressure in the system increased. However, a majority of them also thought that the enclosed air did not exert any pressure if the syringe was not being compressed. In other words, children tend to associate pressure with movement; in the absence of detectable movement, they assume that there is also no pressure (Glough & Driver, 1985). Similarly, Séré (1982), found that 11- to 13-year-old children explained the movement of air in relation to another movement. Consequently if a system was at equilibrium, they believed that there were no forces being exerted. Séré (1982) concluded,

*Children lack the knowledge of atmospheric pressure as a state of reference in order to understand that air—even when immobile—exists, is present, and acts. This state of reference is needed to recognize the effects of pressure and to attribute a pressure to any quantity of air, even when it is not in movement.* (p. 308)

The unobvious and undetectable nature of “stationary” air explains why students are not always aware that pressure contributes to an effect. Therefore, they turn to more obvious and concrete but inaccurate explanations (Kariotoglou & Psillos, 1993).

A second challenge that inhibits students’ understanding of pressure-related concepts is that they tend to reason linearly rather than systemically and relationally (Basca & Grotzer, 2001). Children often attribute causes and effects to a linear, unidirectional model, instead of considering more complex variables, and the relationships between the variables, that might offer a better model of the scientific phenomenon (Grotzer & Twitwiler, 2014). For example, Glough and Driver (1985) found that 12- to 16-year-old children explained drinking through a straw as simply pulling or sucking rather than the result of a pressure differential between air pressure inside and outside of the straw. Oftentimes, relational causality, which refers to the interaction between causes and effects, more accurately explains scientific phenomena. As depicted in the second image of Figure 1, the cause of the floating ball is the result of the interaction of and relationship between the two air pressures—not one (or two) pressures acting disparately. Although a linear model is more conspicuous and accessible, it often does not fully explain the complexity of the phenomenon (Basca & Grotzer, 2001).

These prevailing misconceptions and challenges that prevent children and adults from accurately understanding pressure-related topics in the physical world motivate our study. Because Bernoulli’s principle is illustrated in museums all over the world yet is a conceptually challenging topic to grasp, we hypothesized that the addition of AR could help visitors build better knowledge of the science behind the floating ball. In the following section, we describe previous studies of AR in science learning environments that show promising evidence to support its use in improving science content learning.
Augmented reality to scaffold science learning

Augmented reality (AR) technologies have been highlighted for their enormous potential to enable people to construct new understanding (New Media Consortium, 2014). By layering digital displays (known as “digital augmentations”) over real-world environments, the hybrid display of phenomena provides scaffolds for users to experience and perceive virtual elements as part of their present world (New Media Consortium, 2014; Kirkley & Kirkley, 2004). In so doing, the augmentations help users explore aspects of the world in more concrete ways than might otherwise be possible (Yoon & Wang, 2014).

This potential to augment users’ interactions, engagement, and experiences has revealed numerous affordances of AR for science learning. These include supporting students’ scientific spatial ability, by (a) allowing them to manipulate and learn content in three-dimensional perspectives (Kerawalla, Luckin, Seljeflot, & Woolard, 2006; Martín-Gutiérrez et al., 2010); (b) engaging them in scientific inquiry by encouraging them to make observations, ask questions, collaborate with others, and investigate and interpret data (e.g., Dunleavy et al., 2009; Rosenbaum, Klopfer, & Perry, 2007; Squire & Jan, 2007; Squire & Klopfer, 2007); and (c) enhancing their conceptual understanding by enabling them to visualize invisible or abstract concepts or events (e.g., Clark, Dunser, & Grasset, 2011; Dunleavy et al., 2009; Dunleavy, 2014). For further descriptions of the features and affordances of AR for educational purposes, see Cheng and Tsai (2013), Wu, Lee, Chang, and Liang (2013), and Dunleavy and Dede (2014). These studies demonstrate that, compared with traditional teaching methods, students who use AR applications tend to demonstrate higher academic achievement levels (Ibañez, Di Serio, Bell, Lewenstein, Shouse, & Feder, 2007), and these technologies are essential to support learning. From conveying spatial information and their children. Szymanski and colleagues (2008) revealed that electronic guidebooks have been modified to include digital augmentations. These devices were selected by museum staff because of their potential to augment users’ interactions, engagement, and experiences has revealed numerous affordances of AR for science learning. These include supporting students’ scientific spatial ability, by (a) allowing them to manipulate and learn content in three-dimensional perspectives (Kerawalla, Luckin, Seljeflot, & Woolard, 2006; Martín-Gutiérrez et al., 2010); (b) engaging them in scientific inquiry by encouraging them to make observations, ask questions, collaborate with others, and investigate and interpret data (e.g., Dunleavy et al., 2009; Rosenbaum, Klopfer, & Perry, 2007; Squire & Jan, 2007; Squire & Klopfer, 2007); and (c) enhancing their conceptual understanding by enabling them to visualize invisible or abstract concepts or events (e.g., Clark, Dunser, & Grasset, 2011; Dunleavy et al., 2009; Dunleavy, 2014). For further descriptions of the features and affordances of AR for educational purposes, see Cheng and Tsai (2013), Wu, Lee, Chang, and Liang (2013), and Dunleavy and Dede (2014). These studies demonstrate that, compared with traditional teaching methods, students who use AR applications tend to demonstrate higher academic achievement levels (Ibañez, Di Serio, Bell, Lewenstein, Shouse, & Feder, 2007), and these technologies are essential to support learning. From conveying spatial information and their children. Szymanski and colleagues (2008) revealed that electronic guidebooks have been modified to include digital augmentations. These devices were selected by museum staff because of their

AR technology is also starting to slowly extend into museum spaces. However, as most of these technologies are prototypes and still in the development stages, research on their use in museums is largely concerned with their design, evaluation, and usability (Bell, Lewenstein, Shouse, & Feder, 2009). Some studies have investigated the development of guidebooks to support visitors’ navigation of AR displays and their interactions with the displays throughout the museum (e.g., Damala, Cubaud, Batino, Houlier, & Marchal, 2008; Szymanski et al., 2008), while others have studied the technological design, architecture, and implementation of an AR system (e.g., Koleva et al., 2009; Wojciechowski, Walczak, White, & Cellary, 2004). Although these studies do not specifically examine the impacts on visitor learning, they do offer important insight into the general effects AR has on visitors’ behavior. For instance, Asai, Sugimoto, and Billinghurst (2010) reported that an AR lunar surface navigation system implemented at a science museum exhibit encouraged more collaborative interactions between parents and their children. Szymanski and colleagues (2008) revealed that electronic guidebooks increased visitors’ exploration of the objects being augmented, and Hall and Bannon (2006) demonstrated that children’s engagement and interest increased when they interacted with several museum artifacts that were augmented.

There are only a few studies that look at museum visitors’ knowledge and use of AR. Chang et al. (2014) investigated college students’ appreciation of art by comparing the use of an AR enhanced guide, an audio guide, or no guide. Students who experienced the art museum through the AR enhanced guide showed greater art appreciation compared to the audio and non-guide experience. The behavior and amount of time with the paintings was not significantly different between the audio and AR guided students. The AR guide was credited with having more easily digestible information compared to the audio guide due to the use of visuals. Similarly, Sommerauer and Müller (2014) explored how AR contributed to visitors’ mathematics knowledge in a museum mathematics exhibition. They found that visitors who interacted with the AR enhanced exhibit performed significantly better on knowledge acquisition and retention tests.

These studies demonstrate that AR has the potential to support learning. From conveying spatial information about scientific elements essential to understanding and visualizing phenomena to increasing collaboration and engagement among its users, AR technology offers promise for transforming science learning. However, particularly for informal environments such as science museums, more empirical research is needed to determine whether and how AR supports visitors’ conceptual understanding of science ideas.

ARIEL studies in a science museum

Building on the research described in the previous section, over the last several years we have used the ARIEL project to investigate how augmented reality and various forms of learning scaffolds can improve visitors’ scientific knowledge in an informal science museum setting. To date, three pre-existing exhibit devices have been modified to include digital augmentations. These devices were selected by museum staff because of their
prevalence in science museums and centers worldwide. The first device, “Be The Path,” was augmented to show the flow of electricity when visitors completed an open circuit with their bodies. The second device, “Magnetic Maps,” was augmented to visualize the magnetic field surrounding two bar magnets. And the third device, “Bernoulli Blower” (depicted in Figure 1), was augmented to feature the interactions between two types of air to keep a plastic ball afloat. Jonassen and colleagues (1994) proposed that when investigating the role of media in student learning we should examine the process of learning first, then the role that context plays in understanding the kinds of cognitive tools and their affordances needed to support learning. The ARIEL project has conducted research examining both the learning afforded by AR in the context of an informal learning environment. Our previous research has shown that learning is largely influenced by collaboration among peers while using the AR device (Yoon, Elinich, Wang, Steinmeier, & Van Schooneveld, 2012b), all the while preserving core aspects of informal participation, such as self-directed experimentation (Yoon et al., 2013). In terms of the cognitive tools, results from experiments with our first two augmented devices demonstrate that AR can increase conceptual (content) understanding (Yoon et al., 2012a) and cognitive (theorizing) skills (Yoon et al., 2012b). For understanding the affordances of the media we have shown that learning is supported through the device’s dynamic visualization capabilities (Yoon & Wang, 2014).

Figure 1. Images of the Bernoulli blower device with digital augmentation
It is important to note that the ARIEL studies take place in a science museum where learning is characterized by free choice, individual motivation, and open-ended playful exploration (Bell, Lewenstein, Shouse, & Feder, 2009). In this informal learning environment there is no set curriculum, there is no instructor and the only change to the learning environment is the addition of AR to the Bernoulli Blower. Here the instructional method is intertwined with the media of AR. It is this change in instruction enabled by the dynamic visualization of hidden information in real time that this particular study aims to explore.

In this study, we examine how the digital augmentations in “Bernoulli Blower” can serve as a scaffold for learning about Bernoulli’s principle. Briefly, the exhibit features a physical plastic ball that is able to float in midair because it is caught between the fast moving air coming from a blower attached to the exhibit and the slow moving air in the room. The digital augmentation is produced on a screen that depicts the fast moving air through arrows that point diagonally up and curve around a real time image of the physical plastic ball. At the same time the screen displays the slow-moving air from the room by depicting shorter arrows that point in and at the real-time image of the plastic ball. Although the normal room air moves at a lower speed than the faster moving blown air, the room air exerts greater pressure on the ball and is therefore able to keep the ball floating in the stream of fast-moving air instead of being blown away. Thus the speed and pressure of flowing air are inversely proportional.

Methods

Participants and context

This study was conducted at a large, well-established science museum in a northeastern U.S. city. The students who participated in this study were selected by their teachers, who themselves responded to a mass email invitation sent to middle school (6th to 8th grade; 11 years old to 14 years old) science teachers in the surrounding area. In total, 58 students (41% male, 59% female) from five schools (three charter, two community public) participated (see Table 1 for other demographic data). We specifically targeted students in this grade band because the concept of air pressure is first introduced in 5th grade in our state’s standards; therefore, all students would have some prior knowledge of the science concepts illustrated by the device. This study was embedded within an all-day school field trip to the museum, and participating students were given free general admission to the exhibits. The total amount of time it took to participate in the research was approximately 1 hour.

<table>
<thead>
<tr>
<th>School</th>
<th>Economically disadvantaged</th>
<th>Percent non-white</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public A</td>
<td>75%</td>
<td>31%</td>
</tr>
<tr>
<td>Public B</td>
<td>83%</td>
<td>84%</td>
</tr>
<tr>
<td>Charter A</td>
<td>84%</td>
<td>99%</td>
</tr>
<tr>
<td>Charter B</td>
<td>82%</td>
<td>100%</td>
</tr>
<tr>
<td>Charter C</td>
<td>80%</td>
<td>100%</td>
</tr>
</tbody>
</table>

On the day before the students’ field trip, researchers went to the schools to collect consent forms and to administer pre-intervention surveys of students’ knowledge of Bernoulli’s principle. On the day of the field trip, each chaperoned group (assigned by the teacher, with roughly nine students in each group) was given a specific time to report to the research area, a space commonly used for museum workshops and classes. (Outside of this time slot, students were free to explore the museum per their teacher’s instructions.) When students arrived at the research area, they were randomly assigned to one of two conditions: the non-AR condition (device with no digital augmentation) or the AR condition 2 (device with digital augmentation). In groups of three, students were invited to the research area and shown the device. Depending on the condition that the students were assigned to, the computer screen, which displayed the augmentations, would either be turned on (AR condition) or off (non-AR condition) and the red ball would be lying on the table. The students were told “see if you can make the red ball float” and asked to play with it as if they had found it on the museum floor. After students signaled that they were finished, they were individually asked a set of interview questions about their experience with the device. Their responses were audio recorded and later transcribed. The day after the field trip, researchers went back to the schools to administer post-intervention surveys of student knowledge. In total, the non-AR condition had 29 students (55% female, 45% male, 55% 7th graders, 45% 8th graders) who spent on average 11 minutes and 43 seconds interacting with the device. The AR condition had 29 students (62% female, 38% male, 66% 7th graders, 34% 8th graders) who spent on average 9 minutes and 41 seconds interacting with the device.
Data sources and analyses

Two qualitative data sets were collected, coded, and analyzed to determine how AR impacted students’ conceptual knowledge of Bernoulli’s principle.

Pre- and post-intervention surveys of student knowledge

The pre- and post-intervention surveys consisted of four multiple-choice (MC) questions and one open-ended (OE) response question. These questions were constructed by a team of researchers and are modeled on similar questions found in middle school science textbooks. Three of the MC questions could be considered near-transfer questions, as the correct answers could be directly accessed from the exhibit device itself. The fourth MC question could be considered a far-transfer question, as it asked students to select a real-world situation that illustrated Bernoulli’s principle. The OE response question depicted a similarly constructed device using common household materials and asked, Why do you think the plastic ball floats in the stream of fast-moving air? The complete intervention survey can be found in the Appendix. Students’ responses were coded using a previously validated categorization manual on a six-point Likert scale ranging from limited understanding (1) to complete understanding (6) (Wang, 2014). Refer to Table 2 for a description of the levels of understanding. Two analyses of covariance (ANCOVAs) were separately conducted for the MC responses and the OE responses. For the MC responses, the independent variable in the ANCOVA was the condition students participated in and the dependent variable was the students’ post-intervention MC scores. The covariate was the students’ pre-intervention MC scores. Similarly for the ANCOVA on the OE responses, the independent variable was the students’ condition and the dependent variable was the students’ post-intervention OE responses. The covariate was the students’ pre-intervention OE scores.

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Sample response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Little Understanding</td>
<td>Identifies the air from the blower as the only factor making the ball float. May discuss various aspects of the air that contribute to the phenomenon, but the focus is on the air itself.</td>
<td>The air pressure, if you have too much air pressure, it’ll just push it away. But if you have just the right amount right in the line of symmetry, it will stay right in place.</td>
</tr>
<tr>
<td>2 – Emergent Understanding</td>
<td>Considers features of the ball that may impact its ability to float. Discussion of the air from the blower may or may not be expressly stated.</td>
<td>It was able to float because it was light and it’s plastic so it has air already in it. Since it’s light enough, the air pushing on it won’t make it move around because it’s not just solid. It has enough air in it to make it move.</td>
</tr>
<tr>
<td>3 – Partial Understanding</td>
<td>Acknowledges other sources of forces—that are not associated with the air from the blower—that are involved, such as gravity and the normal air in the room. May describe how the forces interact (e.g., pushing up/down) with each other to keep the ball floating.</td>
<td>The air, the actual air, is pushing down I think. And the one from the tube is going up so then it’s making it float cause it was pushing on the sides.</td>
</tr>
<tr>
<td>4 – Basic Understanding</td>
<td>Begins to consider that the ball floats because of differences in the pressure/force of the various airs involved. These differences may be described in the amount of pressure/force exerted on the ball or simply identified as high(er) or low(er) pressure. Although they may incorrectly describe these differences, they acknowledge that the interaction of the varying amounts of pressures is integral to the phenomenon. (Note: Simply acknowledging that pressures are involved is not enough).</td>
<td>The air was pushing it up so it would stay away from the tube and the air was also going around it so it could stay stabilized. And also like, slow moving air could stabilize it too...but it wouldn’t push it that far. It would just like…it wouldn’t make a big impact.</td>
</tr>
<tr>
<td>5 – Adequate Understanding</td>
<td>Considers both the speed AND pressure of the airs involved understanding there is a relationship between the speeds and pressures, though the relationship may be incomplete or inaccurate. May only address the speed and pressure of just one kind of air.</td>
<td>So the air from the tube is pushing it up and the air around it’s pushing it in to make sure that it doesn’t fall off the air flow that’s keeping it up...The slow air has the lower pressure and the fast air has the higher pressure.</td>
</tr>
</tbody>
</table>
6 – Complete Understanding

Understands that the ball floats because the high pressure from the slow-moving room air keeps the ball in the low-pressure, fast-moving air stream. Both kinds of airs must be explicitly addressed. These responses cannot contain any misconceptions.

The faster air was pushing it up and the slower air had more…pressure so it was pushing it up and down and keeping it in the same spot so it couldn’t really go out of the air circle. (About the slow air) It’s just like…like not moving air has higher pressure than really fast air.

**Interviews**

Post-intervention interviews were administered to understand how students interpreted the concept illustrated by the device. Interview questions included the following:

*What did you learn from playing with this device?*

*How did you learn this?*

For the AR condition students we asked a follow up question that probed whether they felt the AR supported their learning: *Do you think you would have learned this without the digital augmentations?*

Responses were qualitatively mined both for content understanding (using the same categorization manual that we used for the OE post-intervention survey questions) and for the affordances of AR as a visualization tool (Yoon & Wang, 2014).

For both the OE survey and interview data sets, two graduate students were trained on the categorization manual and scored 20% of the OE survey data (12 out of 59), yielding 83% agreement (on 10 out of 12 responses). The discrepancies on codes were negotiated until one code was assigned. After this step, one researcher coded the rest of the OE survey and interview data.

**Results**

Table 3 shows that students in the AR condition scored significantly higher on the MC portion of the knowledge survey than the non-AR students, $F(1, 55) = 8.600$, $p = .005$, effect size (Cohen’s $d$) = .802. However, the difference in means between conditions was not found to be significant for the OE responses on the knowledge survey, $F(1, 55) = 2.679$, $p = .107$.

<table>
<thead>
<tr>
<th>Student knowledge</th>
<th>Mean Pre (SD)</th>
<th>Mean Post (SD)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC Condition 1</td>
<td>1.517 (.726)</td>
<td>2.000 (.598)</td>
<td>29</td>
</tr>
<tr>
<td>MC Condition 2</td>
<td>1.793 (.789)</td>
<td>2.517 (.688)</td>
<td>29</td>
</tr>
<tr>
<td>OE Condition 1</td>
<td>1.862 (.785)</td>
<td>1.931 (1.193)</td>
<td>29</td>
</tr>
<tr>
<td>OE Condition 2</td>
<td>1.759 (.773)</td>
<td>2.276 (1.750)</td>
<td>29</td>
</tr>
<tr>
<td>Interview Condition 1</td>
<td>n/a</td>
<td>2.660 (.550)</td>
<td>29</td>
</tr>
<tr>
<td>Interview Condition 2</td>
<td>n/a</td>
<td>3.100 (.770)</td>
<td>29</td>
</tr>
</tbody>
</table>

Regarding students’ interview responses, an independent samples $t$-test showed that there was a significant difference between the knowledge scores of students in the non-AR condition and students in the AR condition, $t(56) = -2.543$, $p = .014$, effect size (Cohen’s $d$) = .658 where more AR condition students scored in the higher levels of understanding. In the non-AR condition, 38% of the students had a Level 2 understanding (defined as describing observations or listing objects or concepts presented), and 59% had a Level 3 understanding (defined as identifying a relationship between two of three variables – air speed, air pressure, and the floating ball). In contrast, in the AR condition (device with digital augmentation), although there was a similar frequency of students with a Level 3 understanding, only 17% had a Level 2 understanding, and 21% reached a Level 4 (defined as identifying the involvement of both air speed and pressure) or 5 (defined as recognizing a relationship between varying air speeds and pressures) understanding.

Collectively, these results suggest that the digital augmentation had a positive effect on students’ content knowledge. A perusal of student responses illustrates how the AR influenced their understanding. For example, one student (ID6) in the AR condition who scored a Level 5, said,
It helped you see the air currents that [were] coming from the tube and it helped you see the high pressure air that was coming in from below and above. If air is moving quickly, it has low pressure. If it's moving slowly, it has high pressure.

This student went on to explain that the activity was different from how they normally learned in school because of the screen and the display where she could “experience what it was instead of reading about it in a textbook.” She and two other girls also “tried to play a game” in which they had to “get the ball to move around without completely cutting off the air current.” Here we can see that the student was able to build an accurate understanding of the phenomena while at the same time engaging in self-directed experimentation, which is an archetypal characteristic of informal participation.

Other student responses illustrated how the AR acted as a scaffold for more accurate understanding of the phenomenon. Students commented on the affordance of visualizing scientific details:

...what kind of air like how many, how much pressure do you need to put on it or what kind of pressure do you need to put on it slower, quicker, faster. (ID186)

Students also commented on the affordance of visualizing hidden information:

That the ball was being caught in the air pressure between the outside and what’s coming from the tube... In school, we don’t really have the, I would say the visual learning of it. We just picture it in our minds. (ID17)

These responses demonstrate the affordances of the AR as well as students’ ability to acquire an accurate understanding of how Bernoulli’s principle works in the brief time they were exposed to the exhibit during their museum visit.

Discussion

In this study, we hypothesized that augmented reality, because it enables the visualization of typically invisible causal mechanisms that underlie complex phenomena, could be used as an effective scaffold to help visitors learn about challenging scientific content in a museum. Through the multiple-choice portion of the knowledge survey and interviews, our results indicate that students in the AR condition significantly improved in their understanding of Bernoulli’s principle and showed greater gains compared with students in the non-AR condition. That the results from the open-ended response portion of the knowledge survey did not show a significant difference between conditions may be attributed to the fact that those kinds of surveys draw on limited information while interviews provide the opportunity for students to be probed for deeper understanding. Student interviews also showed that the AR served as a valuable learning scaffold by enabling students to visualize scientific details, recognize and make sense of hidden information, and gain a more accurate understanding of the science. Elsewhere we have documented similar dynamic visualization affordances of another, conceptually less-challenging AR device in our ARIEL series of experiments (Yoon & Wang, 2014), which points to the continuing validity and reliability of these results.

These results also support the viability of AR in challenging misconceptions of pressure-related concepts concerning the undetectable nature of air movement (Kariotoglou & Psillos, 1993) and students’ inability to recognize complex relationships between scientific variables (Basca & Grotzer, 2001). From the interviews, we can see clear evidence of students’ reasoning accurately about the inverse relationship between fluid speed and pressure in Bernoulli’s principle (Hewitt, 2004)—they recognized that slower moving air has higher pressure and faster moving air has lower pressure. Students were also able to articulate how the two kinds of air worked together to make the ball float, which attests to growth in understanding of the complex relationship of variables (Basca & Grotzer, 2001). Furthermore, some students were able to identify the non-obvious influence of the surrounding air pressure, which they would not have understood without the AR.

Our findings in this most recent study in our ARIEL series investigating the use of AR devices in museums (e.g., Yoon et al., 2012a) convince us that AR can be used to support learning in informal environments through specific scaffolds. We have shown in this study that AR not only supports learning of science content but can also support learning of very challenging science content during brief periods of exploration. We acknowledge that our results, though positive, revealed only modest conceptual gains, with participants’ improving on average less than a level on the knowledge survey and reaching just above partial understanding in the interviews. Looking forward, we are considering additional ways to scaffold the experience to induce greater learning while at the same time preserving the informal experience. Designing for content understanding will inevitably require increased and multiple scaffolds (Reiser & Tabak, 2014), some of which are already common practices within
museums, such as grouping exhibits into clusters based upon conceptually related content. In our future work, we will investigate various aspects of exhibit design, in addition to what we have learned about AR devices in museums, to understand how content learning is best facilitated.

Acknowledgements

This research was funded by NSF DRL Grant #0741659.

References


Appendix

Knowledge Test

1

What is the relationship between the speed and pressure of moving air?

a. The faster the speed is, the higher the pressure.
b. The slower the speed is, the lower the pressure.
c. The faster the speed is, the lower the pressure.
d. Speed and pressure are not related.

2

A) When the air blower begins to blow, what will happen to the ball?
   a. It will blow away and fall to the ground.
   b. It will stay floating in the stream of fast-moving air.
   c. It will float for a few seconds and then blow away.

B) Which air puts MORE pressure on the ball?
   a. The fast-moving air from the blower (solid line).
   b. The slow-moving air around the ball (dotted line).
   c. Both put the same amount of pressure (solid & dotted lines).
Look at the picture.

The electric hair dryer is blowing a stream of fast-moving air upwards.

A small, lightweight, plastic ball is floating in the stream of fast-moving air. It will continue to float there until the hair dryer stops blowing.

Why do you think the plastic ball floats in the stream of fast-moving air?

Which of the following real-world scenarios illustrates the concept you just learned? (Circle all that apply).

a. running a race
   b. paper lifting off the desk due to a gust of wind
   c. climbing a mountain
   d. ejecting out of a plane
Online Research Behaviors of Engineering Graduate Students in Taiwan

Ying-Hsueh Cheng1* and Chin-Chung Tsai2

1National Pingtung University of Science and Technology, Taiwan // 2National Taiwan University of Science and Technology, Taiwan // sherrycheng85@gmail.com // cctsai@mail.ntust.edu.tw

*Corresponding author

(Submitted September 9, 2015; Revised December 26, 2015; Accepted September 30, 2016)

ABSTRACT

Previous studies have examined the online research behaviors of graduate students in terms of how they seek and retrieve research-related information on the Web across diverse disciplines. However, few have focused on graduate students’ searching activities, and particularly for their research tasks. Drawing on Kuiper, Volman, and Terwel’s (2008) three aspects of web literacy skills (searching, reading, and evaluating), this qualitative study aims to better understand a group of graduate engineering students’ searching, reading, and evaluating processes for research purposes. Through in-depth interviews and the think-aloud protocol, we compared the strategies employed by 22 Taiwanese graduate engineering students. The results showed that the students’ online research behaviors included seeking and obtaining, reading and interpreting, and assessing and evaluating sources. The findings suggest that specialized training for preparing novice researchers to critically evaluate relevant information or scholarly work to fulfill their research purposes is needed. Implications for enhancing the information literacy of engineering students are discussed.

Keywords

Graduate students, Online research behavior, Information searching, Strategies, Engineering

Introduction

With the rapid development of digital technologies, a variety of sources ranging from Google/Google Scholar to library databases such as the Web of Science and Scopus play crucial roles in graduate students’ information seeking processes when they are confronted with certain research tasks (Du & Evans, 2011; Rempel, 2010). Using search engines and library databases as entry points, graduate students need to specify keywords for searches, browse search results, chase references, assess and extract relevant information from sources, and synthesize information for specific research purposes (George, Bright, Hurlbert, Linke, St Clair, & Stein, 2006). However, for most graduate students, mastering these skills can be difficult because they might lack the ability to critically process and evaluate academic literature and thus fail to effectively retrieve information (Ismail & Kareem, 2011; Wu & Tsai, 2007).

How graduate students actually handle and use online information for research purposes has received increasing attention (Al-Muomen, Morris, & Maynard, 2012; Catalano, 2013; Kerins, Madden, & Fulton, 2004). During the past decades, several studies have examined graduate students’ information searching behaviors in diverse disciplines, including law (Makri, Blandford, & Cox, 2008), the humanities (Barrett, 2005; Bronstein, 2007), physics and astronomy (Jamali & Nicholas, 2010), and basic and medical sciences (Hemminger, Lu, Vaughan, & Adams, 2007; Liang & Tsai, 2009). However, few studies have been undertaken to investigate graduate engineering students’ searching behaviors. As Harrison (2009) pointed out, “student engineers are almost universally not technological novices and they know their way around the online world, but being able to work out which algorithm a search engine uses to produce results is not the same thing as being able to work out how to get useful results from that search engine quickly and efficiently” (p. 68). Harrison highlights the challenge of retrieving useful information with the least effort for engineering students. This implies the need to examine engineering students’ searching practices and strategies in this rapidly changing digital era. This study was thus conducted to address these issues that are often not taught in the classroom.

Literature review

Graduate students’ online research behaviors

During the past decades, terms such as information searching or seeking have been used to describe how Web users find and select appropriate information resources to increase their knowledge (Ellis, 1993; Lee & Cho, 2011; Wang, Liang, & Tsai, 2014). Considering this study, we use the term online research behaviors, coined by
Biddix, Chung, and Park (2011), to indicate the intricate information seeking processes academic novices such as graduate students actually engage in.

For some time now, researchers have been interested in understanding graduate students’ online research behaviors relating to information searching activities and strategies (Ismail & Kareem, 2011; Makri et al., 2008; Vibert, Rouet, Ros, Ramond, & Deshouilleres, 2007). Some have conducted large-scale studies with the intention of categorizing students’ search behaviors across diverse disciplines (Du & Evans, 2011; Ellis, Cox & Hall, 1993; George et al., 2006; Rempel, 2010). For example, examining doctoral students’ searching patterns across natural and social science disciplines, Du and Evans (2011) found that these students adopted various strategies, including interacting with multiple search systems (search engines, online databases, specific websites), relying on popular search engines such as Google or Google Scholar, modifying keywords or search queries (using the Boolean operators: and, or, not) and other operators (* “” +), as well as looking for two or more topics simultaneously.

In contrast, others have adopted a discipline-based approach (Grafstein, 2002; Jamali & Nicolas, 2010). Rather than looking at the generic skills related to the general process of retrieving and evaluating information, these researchers emphasize the need to examine the skills required for acquiring knowledge or conducting research in a specific subject area (Talja, Vakkari, Fry, & Wouters, 2007). Following this line of investigation, an increasing number of researchers have started to examine graduate students’ searching practices in distinct disciplines. For example, Barrett (2005) explored graduate humanities students’ searching practices and found that they used Google to find general information on a topic, and other techniques such as citation chasing, identifying primary sources to validate their theories and hypotheses, being guided by interpersonal contacts, constantly reading in a subject area, and dealing with time pressure. In a study related to doctoral neuroscience students’ information searching, Vibert et al. (2007) reported that these students conducted searches to find out about experimental techniques, understand “the great debate,” and keep track of publications in their field. Exploring the relationship between disciplinary differences and searching behaviors, Jamali and Nicholas (2010) found that the online research behaviors of doctoral students from physics and astronomy tended to be interdisciplinary and involved reliance on the literature of other subject areas due to the wide-ranging information in their own disciplines.

Although these studies shed light on discipline-specific searching behaviors, research on graduate engineering students’ online research activities is still underexplored. Relevant to this study, Ismail and Kareem (2011) investigated the information seeking of master’s students enrolled in a computer science and information technology program in Malaysia. The authors found that these students faced difficulties in finding relevant information related to their research. Due to the lack of specific scholarly tools at the early stage of their research work to answer some important questions, the students relied on Google or Google Scholar for searching, but were overloaded with massive amounts of information. Ismail and Kareem’s study called attention to novice engineering researchers’ struggle to obtain the required information, and highlighted the importance of facilitating the acceleration of their searches for research-related information.

This study thus aimed to explore Taiwanese engineering graduate students’ online research behaviors with a focus on three issues: (1) How do they seek and obtain online information for research purposes? (2) How do they read and interpret sources? and (3) How do they assess and evaluate those sources for their research tasks?

Theoretical framework

To answer these three questions, we drew on Kuiper, Volman and Terwel’s (2008) theoretical framework. Kuiper and his colleagues (2008) proposed three aspects of Web literacy skills: searching, reading, and evaluating. Searching skills include the ability to use search engines or library databases, define appropriate keywords, and use search options within a specific website. Reading skills involve the ability to sort through the masses of information and make decisions on what to use and what to neglect. Evaluating skills comprise the ability to assess the relevance, reliability, and authority of online information. According to Kuiper et al. (2008), these three skills overlap and are mutually connected. They argued that “the use of the Web always involves all skills; for example, searching for Web information always involves reading and evaluating skills” (p. 690). Although Kuiper et al.’s (2008) framework was developed for teaching 5th graders how to search, read, and evaluate Web information, it can be extended to the graduate school context. We argue that this framework can be a useful tool for interpreting graduate students’ searching behaviors related to their research tasks. Therefore, the present study aimed to investigate how graduate students search for, read, and evaluate information sources for their research purposes with the application of Kuiper et al.’s (2008) framework. It is hoped that the findings of this study can contribute to information literacy education for engineering students.
Methods

Participants

In order to understand graduate engineering students’ information searching in Web-based environments, 15 master’s students and seven doctoral students from two national universities in Taiwan were individually interviewed (see Table 1 below for details). The students’ research-related searching experiences ranged from one to seven years. Their areas of specialization varied from material science and engineering, automation and control, chemical engineering, electronic and computer engineering, electro-optical engineering to applied science and technology. They conducted academic literature searching for a variety of writing purposes (course assignments, grant projects, conference papers, journal articles, and theses/dissertations) and sometimes for experimental purposes, when they searched for the methodology details of other studies.

Table 1. Participants’ backgrounds and publication experience (n = 22)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Number</th>
<th>Average age</th>
<th>Gender</th>
<th>Years</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master’s students (MS)</td>
<td>15</td>
<td>24</td>
<td>11 males, 4 females</td>
<td>Two</td>
<td>Three published 1-4 journal papers</td>
</tr>
<tr>
<td>Ph.D. students (Ph.D.)</td>
<td>7</td>
<td>27</td>
<td>5 males, 2 females</td>
<td>Two to Four</td>
<td>Five published 1-8 conference papers</td>
</tr>
</tbody>
</table>

Data collection

The data set included two parts: data from semi-structured retrospective interviews and from the concurrent think-aloud protocol. Audio recordings were collected from the interviews, and screen recordings were gathered from the think-aloud protocol.

Interview

One individual face-to-face interview (30-40 minutes) was conducted by the first author with each participant. The interviews were conducted in Chinese, a native language shared by the researcher and the participants. The first part mainly explored each participant’s history of searching for and using online information for research-related assignments since college. The second part focused on how they were guided or self-taught to search, read, and evaluate Web information for research projects involved in the graduate program. These questions included: “What search engines or library databases do you usually turn to? How do you use them? How do you select articles or information for your research purposes? Who do you turn to if you have questions about the search results or research direction?” The second part asked the participants to conduct literature searching on a laptop computer and concurrently explain their evaluation standards and strategies for selecting relevant literature. All of the interviews were transcribed and selectively translated for inclusion in this paper. The translated data were verified by the second author.

Think-aloud video

Following the interview, the think-aloud protocol was employed to compare the statements and actions of the participants (Van Someren, Barnard, & Sandberg, 1994) who were asked to demonstrate what online library databases or search engines they commonly used and to explain orally how they searched for, read, and evaluated sources based on their recent experiences. The Chinese version of a free software package called HyperCam was used for screen recording. Due to technological issues, one screen recording was lost and thus 21 were used for comparison with the interviews. Observation was conducted during the think-aloud protocol. The first author took notes of the challenges and strategies the participants used for locating, using, and assessing online sources.

Data analysis

As indicated above, the analyses were mainly guided by Kuiper et al.’s (2008) theoretical framework: Searching skills, reading skills, and evaluating skills.
Overall, three kinds of coding strategies were used for analysis: descriptive codes, interpretative codes, and pattern codes (Miles & Huberman, 1994). Descriptive codes are processed in a straightforward way, entailing little interpretation. Interpretive codes require more interpretation of data sources by the researcher. Pattern codes are inferential and explanatory notes, signaling themes or regularities that occur.

Field notes and screen recordings were analyzed with the emphasis on the ways the students sought and obtained information, read online sources, and evaluated the materials. These data sources were triangulated with the interviews for comparison and contrast.

Findings

The participants illustrated three aspects of strategies for online research, including searching, reading, and evaluating. While the first two are lower-level skills, the last is a higher-level strategy for critically assessing the usefulness and accuracy of online information. The participants used the three types of strategies recursively and discursively during their research processes. In the following, we delineate how they used these strategies.

Seeking and obtaining information

The participants reported four strategies they used for seeking and obtaining online information. These strategies included: using search engines and library databases, using and modifying keywords, tracking references, and networking for retrieving articles or materials.

Using search engines and library databases

Of the 22 participants, 14 revealed that they turned to the library databases before Google. They relied on major databases in the sciences such as IEEE, Nature, Scopus, ScienceDirect, the Web of Knowledge, and the Web of Science. Some students indicated that, of all the databases, the library ones were more useful because they provided a broad coverage of topics in their area, allowing them to find the journal papers they wanted by specific material or year, and most importantly, these databases collected top-quality journal papers. Others reported that they turned to the library databases before Google or Google Scholar because of the influence of their seniors or advisors who valued sources from these databases. In contrast, seven students reported that they turned to Google or Google Scholar before consulting the library databases. Their purpose was to browse related topics to establish background knowledge so that they could determine what specific journals they could turn to or what products were already available on the market that needed to be modified or invented. According to them, these search engines were useful because they consolidate major databases in the sciences and thus provide sufficient sources for their searches.

Using and modifying keywords

A total of 15 students reported that they used and modified keywords when using the search engines or library databases to find the sources they needed. They began their searches by identifying primary keywords and then continued with other approaches such as adding or changing keywords or using Boolean operators (AND or “ “) for better results. Below are examples:

Researcher: Do you use keywords? How do you use them?
Student 1 (MS): Keywords. OK. For example, if there are A, B, C materials, usually we can’t type in all of them…because that will limit the search results. Usually I type in A first and it works better for searching. You can also type in A and B together, or A and C for classification and that [such use of keywords] has fewer limitations [than typing in several keywords altogether].

Student 2 (MS): Usually I don’t use “and” [between keywords]. I use keywords directly instead. Like I use “sysweld” as a keyword and then a space after that. The method I use is called Tig welding [Tungsten Inert Gas arc Welding]. If I want something to appear in the searches, I add quotation marks around the word “Tig” so that the results will appear with this word.
The participants noted that they learned to use the keywords by referring to the journal articles or through their advisors’ or seniors’ guidance.

**Tracking references**

Nine students reported that they tracked references when they needed to understand certain statements or concepts in depth. For example, Student 16 (MS) noted that she tracked references when a certain statement seemed crucial but was too brief to understand. She would refer to the references for clarification.

When I feel I need to understand...a certain statement that is very brief [in the article] but which looks critical, I would look at the references…and search [the article for more information].

Student 17 (Ph.D.) also discussed his experience while searching for information for a grant project conducted by his advisor. In collaboration with an aerospace center, Student 17 was provided with a few sources by the center and then found more articles based on those original sources.

They [the aerospace center staff] gave us some useful stuff...When I just read it, I found it hard to understand. I saw some…concepts that were worth exploring…from the references…yes…references are very good. References could help expand the resources.

**Networking for retrieving articles or materials**

Ten participants indicated how they networked with others such as peers, friends, or seniors studying or working at other universities to retrieve relevant articles or materials. Student 14 indicated that she reached out to a friend studying at another university for downloading a thesis related to “game theory.”

Student 14 (MS): I obtained the thesis and codes from a classmate of mine…I was learning a new theory called “Game theory.” It is very new and my lab was doing it. My classmate told me one of his seniors used this for his thesis and so I got it from him.

Student 15 reported that he used to contact a friend working as a postdoctoral researcher at another university for retrieving articles from two online journals, *Nature* and *Science*.

Researcher: Some people have [external] connections for using library accounts or passwords to download articles. Do you also have such connections?  
Student 15 (MS): Yes…our school spent a lot of money purchasing some databases…but compared to databases in some [top-ranked] universities, [what we have] is far less complete. There is a senior who graduated from my school and now he’s working as a postdoc in another top university. Sometimes if we couldn’t find a research paper, we asked him for help…like each discipline has its own most influential journals, such as *Nature* and *Science*, these two we don’t have at our school [and so we asked him to search for papers in these journals].

Student 22 (Ph.D.) had a similar experience of borrowing her classmate’s account and password for accessing external library databases and thus was able to download articles she needed from there.

Previously when he [my classmate] was studying for his master’s…now because my classmate graduated, I don’t have his account anymore. He gave me his account and password so I could access the VPN [Virtual Private Network], a kind of network [to access the library system]? Anyway, a website, and [I] got access to it so I could download [what I needed].

Others turned to suppliers of laboratory equipment or conference presenters for obtaining useful information.  
Student 1, a master’s student in material science, mentioned his experience of talking with a supplier of his laboratory equipment. Learning from the supplier, the student knew the importance of tracing back to the original paper in the literature related to his research.

Researcher: How did you know you need to refer to the original paper in the literature?  
Student 1 (MS): One supplier told us…You need to look at references…and trace back to the original.
Student 21, a second-year doctoral student in Optoelectronics, noted that talking with conference presenters can be useful. He indicated,

In the conferences...I go to oral paper sessions and poster sessions...if I see anything related or interesting, I ask the presenters questions...If I am not sure about what they are doing, I bring the topic back home and search online...see if I can download [find] anything.

**Reading and interpreting information**

The participants reported that they tended to skim the title and the abstract and read specific sections in the article to locate related information. Moreover, to develop a better understanding of a research topic, they were aware that they had to read broadly in a subject area.

**Skimming title and abstract**

A total of 11 participants (10 males, 1 female) reported that they tended to skim the title and the abstract to identify if certain information such as research directions, methods, or results were related to their research. The title and the abstract become entry points for the students to evaluate the relatedness of sources and to justify whether they should read the main text further.

Student 13 (MS): I usually look at the title first to see if it’s related to our research. If it is interesting, I would look at the abstract... If it [the abstract] is related to what I am doing..., then I would read closely the experimental procedure, and what kind of analysis was conducted.

Similarly, Student 12 (Ph.D.) noted that he usually looked at the methods and results in the abstract. If the results were positive or unexpected, he would continue to read the paper. However, if the direction was unrelated, he would not continue to do so. By contrast, unfamiliar with these reading strategies, Student 14 (MS) reported that she used to read the full paper from beginning to end and later she realized that she could turn to the abstract first to get a basic idea of the paper.

**Referring to specific article sections**

After reading the abstract, 12 participants mentioned that they read specific article sections such as the introduction \( n = 2 \), methods and results \( n = 9 \), and conclusion \( n = 1 \) for related information. Student 17 (Ph.D.) indicated that he read the introduction to understand the research gaps and contribution of the study. However, one student, Student 8 (MS), indicated that after skimming the abstract, he usually read the conclusion to verify if the direction was what he wanted to pursue. Most students reported that they referred to graphs and experimental details such as parameters, temperature, and frequency because those were often the key information in scientific studies. As was noted by Student 12 (Ph.D.), “Methods play a central role in the article, as do the results. It’s like building on the [previous] methods to advance research...I look at whether the methods were newly developed or similar to what I am doing...I also look at the differences in the time domain and frequency domain in the graph and many other graphs related to the results.” Similarly, Students 16 (MS) and 18 (Ph.D.) both indicated that when downloading a paper, they pay attention to the images or graphs. Student 18 further noted, “I focus on the images because they are easy to understand. After I look at the images, if the content is related to my research, I will read the full paper.” These results suggest that developing the ability to interpret graphs and images is important training for engineering students since key information is often embedded in these data. Lacking such ability might cause difficulty in judging whether the information is related to one’s research purpose.

**Reading broadly in a subject area**

Several students \( n = 8 \) noted that they read broadly and continuously in a subject area for a period of time. Student 6 (MS) located more than 20 articles related to the topic of “uneven brightness” and decided to read 10 of them in detail. Student 12 (Ph.D.) indicated that some articles he downloaded were not directly relevant to his topic. His rationale was that at the beginning of the search stage, he might not be experienced in judging the usefulness of the literature and thus he reads articles from multiple directions to increase his understanding of a
certain topic. Student 21 (Ph.D.) noted, “It’s difficult to understand what it [the topic] is about, if you only dig in a certain direction.” Thus, he reads broadly to gain a basic understanding of a topic and to identify what could be improved.

Assessing and evaluating information

Four evaluation criteria were discussed regarding the usefulness of online sources. These included: relevance, recency, credibility, and authority. How the students used these criteria is illustrated as follows.

Relevance

A total of 13 students evaluated topical relevance based on their laboratory directions and resources. In other words, their proposed research must build on the previous laboratory work. When they conduct online searches, they use terms or keywords related to their laboratory experiments or projects. For example, Student 22 (Ph.D.) noted that for her research to be manageable, she takes into consideration the availability of equipment in her laboratory. Based on this understanding, she types in the word, “antenna” in Scopus in her searches and focuses on the results related to this topic. Likewise, Student 1 (MS) stated: “First of all, I would consider equipment…because I am going to conduct an experiment. For example, if I found a topic, and it used a different processing mode…If the school does not have the equipment, I would not read this article.”

Recency

A total of 14 students pointed out that, influenced by their previous training, they tend to find articles that are published within three to five years. They search for the articles by year and pay close attention to the changes or trends in the field. Student 17 (Ph.D.) noted that he tends to begin his searches from the most recent and then looks back. While he is doing this, new ideas might occur and help him shape his research. This indicates the rapid development and changing nature of scientific research which forces the engineering researchers to update their literature. In contrast, a few students noted that they refer to older sources because of the need to apply specific mechanisms or to use articles with high citation rates.

Credibility

Nine participants mentioned that they value journal articles more than conference papers. Moreover, they value international journal articles more than domestic ones. Some students did not exactly understand why they have to do so, indicating that they just follow their advisor’s suggestions. Others mentioned that international journal articles are more credible and reliable and that these sources become their indicators when they are preparing their manuscripts for submission. The students judge the credibility of sources based on either quantitative indicators (e.g., citation rates or impact factors) or publishers based in larger countries (e.g., Europe or the United States).

Authority

The students revealed distinct opinions regarding whether they would evaluate the sources based on the authority of the author. Seven students noted that they pay attention to a particular author’s or research team’s work because these studies are pioneers of certain projects and thus are well-known in the subject area. In contrast, 10 students reported that they do not use this criterion to judge online information for their research. Some noted that they pay attention to methodology and others look at the relevance of the topic. For example, Student 2 (MS) revealed that he does not track a certain author’s works because not many researchers conduct related works in his area. He would consider himself fortunate if a few sources were found. Such a result implies that when there is insufficient information in a certain subject area, the students tend to adjust their criteria instead of following the authority of the author.

In summary, the students evaluated the usefulness of their sources based on four criteria, namely relevance, recency, credibility, and authority. They learned to use these criteria flexibly and strategically when their research purposes varied.
Discussion and conclusion

The findings correspond with the three aspects of web literacy skills raised by Kuiper et al. (2008): searching, reading, and evaluating. In the following, we offer responses to the research questions and discussion of the findings in relation to the literature.

Seeking and obtaining information

Previous studies reported that due to a lack of familiarity with academic resources, college students tend to turn to Google or Wikipedia before consulting the library (Biddix et al., 2011). However, in this study, we found that most graduate engineering students turned to the library databases before Google. They relied on major databases such as IEEE, Nature, Scopus, ScienceDirect, the Web of Knowledge, and the Web of Science to locate relevant information, confirming results observed in the research practices of experts (Meho & Yang, 2007) and graduate students (Hemminger et al., 2007). Influenced by their advisors and seniors, the graduate engineering students in the current study valued these resources because of the collection of high quality journal articles. With some training, they had learned to find articles by topic or publication year, contradicting the finding that college students rely heavily on non-academic sources from Wikipedia. This finding indicates the importance of preparing undergraduate engineering students to use professional databases for information retrieval. It also highlights the nature of engineering graduate students’ acquisition of information literacy through utilizing academic databases (Messer, Kelly, & Poirier, 2005).

Frequent use of Google or Google Scholar by graduate students has been documented in the literature. Purposes reported include conducting broad searches (Rempel, 2010) and using them as a starting point (Du & Evans, 2011) or to acquire new knowledge (Vibert et al., 2007). A similar result was found in the current study. Seven students indicated that they turned to Google or Google Scholar before the library databases because they consolidated major science databases and thus provided sufficient sources related to their research. This implies that when unfamiliar with a research topic, it is beneficial to begin research using search engines to browse and construct background knowledge and then to use library databases for specific search purposes.

Networking for retrieving articles or materials was frequently reported by the graduate engineering students in this study. This is similar to the results in Hertzum and Pejtersen’s (2000) study. Aligned with previous studies focusing on expert engineers, the finding showed that online engineering research behavior involves dynamic source-seeker connections for retrieving information efficiently (Xu, Tan, & Yang, 2006). This was also indicated by Ellis et al. (1993) regarding science and engineering experts’ consultation of colleagues as a starting point of their research cycle. It is suggested that graduate students initiate contacts with colleagues, librarians, or experts inside or outside their programs to request information.

Reading and interpreting information

This study reveals how graduate engineering students read and interpret information, an issue that is seldom discussed in the literature of information seeking. The participants reported that they select main ideas for reading, including browsing the title first and then the abstract, and referring to specific sections in the article (e.g., methods or conclusion). This corresponds with Tsai and Tsai’s (2003) finding of selecting the main idea strategy regarding how students grasp or summarize the main information provided in each Web page. These findings are important because they involve quick decisions to read online or to download an article for in-depth reading. The engineering students’ practices for online and printed resources are more intertwined and interactive than those observed in Ellis et al.’s study (1993).

Moreover, the current study observed that the engineering graduate students evaluated whether the information was relevant or useful based on the graphs or images in the methods section. This is similar to Aurisicchio, Bracewell, and Wallace’s (2010) finding regarding aerospace designers’ use of drawings.

In addition, it was found that the participants were aware that they had to read broadly in a subject area to develop a comprehensive understanding of a research topic. This finding is consistent with graduate humanities students’ practices (Barrett, 2005) and highlights the importance of building domain knowledge through searching and reading.
Assessing and evaluating information

Four criteria were mentioned by the students when evaluating the usefulness of sources: relevance, recency, credibility, and authority. Differing from the processes of searching and reading, these criteria are considered to be higher-order thinking skills which facilitate the critical differentiation and sorting out of online information. Without this process, online research would not be successful and thus it plays a crucial role in information seeking (Biddix et al., 2011; Currie, Devlin, Emde, & Graves, 2010; Head & Eisenberg, 2009). This finding supports Hofer’s (2004) view regarding the importance of understanding how academic researchers “evaluate sources of knowledge, coordinate theory and evidence, and justify their knowledge assumptions” (p. 51). This reflects the fact that judging, selecting, and synthesizing information for research purposes requires repeated learning and practice. Aligned with Hofer’s perspective, Grafstein (2002) noted that “Given the seductively easy accessiblity of masses of unregulated information, it is imperative that students, from the very beginning of their academic careers, adopt a critical approach to information and develop the ability to evaluate the information they encounter for authenticity, accuracy, credibility, authority, relevance, concealed bias, logical inconsistency, and so on” (p. 199). This implies that novice engineering students can use the above criteria to assess and evaluate information sources for their research tasks.

Limitations and implications

Research limitations and implications

This study has some limitations and thus some implications can be drawn for future research. First, we did not compare differences in the strategies adopted by master’s and doctoral students. Future studies can investigate whether there are any differences between these two groups of students’ searching, reading or evaluating behaviors. Second, in this study, a sample of Taiwanese graduate engineering students enrolled in master’s and doctoral programs was recruited. It is suggested that the theoretical frameworks and the conclusions of the study can be further explored in other student populations. Moreover, in this study, we only recruited 22 participants. Future studies could expand the sample size and include more students from different backgrounds and educational contexts in order to gather a wide range of views. Interviews with those (e.g., librarians, system developers, peers, colleagues, or advisors) who facilitate the participants to conduct their literature searching may be needed in order to triangulate varied perspectives of their searching processes. Adopting mixed methods including surveys and interviews could perhaps gather a broader picture of users’ searching perceptions and behaviors. The findings of this study can serve as the basis for developing survey items to further explore how factors such as gender and different search contexts influence researchers’ searching processes.

Pedagogical implications

This study aimed to investigate the online research behaviors of graduate engineering students that have been underexplored in the past. In addition to the theoretical contributions, this study provides some pedagogical implications for course instructors, advisors, and educational system developers.

For course instructors, it is suggested that they pay attention to engineering graduate students’ difficulties, and provide strategy training to help them retrieve information more efficiently. In our study, graduate engineering students reported that when they just began to conduct lab research, it took them a while to know how to differentiate and locate sources related to their research problems. After consultation with their seniors and advisors, they learned how to identify sources that were relevant, recent, and credible. The ability to judge, select, and synthesize information for specific research purposes is essential. Thus, we believe that offering strategy training can help students to set up criteria to critically assess information.

For advisors, it is crucial to consider resources and services to support graduate engineering students’ information literacy development both online and offline. In our interviews, we found that reading English research papers for some graduate students with inadequate English proficiency might be a big challenge. Several students noted that in their first year of graduate study, their advisors formed study groups which guided them in how to read research papers. They indicated that this process was beneficial because it helped them to comprehend the research articles and know how to track references online. Thus, forming study groups might be especially needed for those who have just begun to work on their lab research.
Aligned with previous studies (Rempel, 2010; Du & Evans, 2011; Vibert et al., 2007), our findings revealed that some students turned to Google or Google Scholar before library databases because these online search engines consolidated major databases in the sciences and thus provided sufficient sources for their searches. We suggest that system developers instruct novice scholars in how to use advanced Google search commands such as modifying search queries or keywords (using the Boolean operators: and, or, not) and other operators (* " +”). It would also be of value if system developers could instruct students in how to find articles by author, topic, or publication year.

References


The “flipped classroom” as a technology-supported pedagogical innovation, has gained prominence worldwide. “WHAT” is the flipped classroom? Although there is no single definition, it is generally characterized by its course structure comprising in-class and out-of-class activities. It uses classroom time for students to actively engage in interactive learning activities, while traditional lectures are delivered out of formal class time with videos, audios, content-rich websites, games and simulations, and the like (Baeplter, Walker, & Driessen, 2014). Based on sound pedagogical and theoretical principles, the “flipped classroom” approach aims at exploiting classroom time and space for appropriately designed interactive learning activities to foster students’ active learning, in which students engage in collaborative and problem-based learning activities to develop higher order thinking skills (Prince, 2004). Such an instructional design intends to have lecture instructions out of class time, thus affords more classroom time to engage students in active learning (e.g., Jungić, Kaur, Mulholland, & Xin, 2015; Sohrabi & Iraj, 2016; Wasserman, Quint, Norris, & Carr, 2015) to bring about paradigm shift where the teacher becomes a “guide on the side” instead of a “sage on the stage” (King, 1993), and students are empowered to explore and solve problems either independently or in groups to achieve their learning goals.

Although a common understanding is reached on “WHAT” the flipped classroom is, there is still a pressing need for studying “HOW” to implement the flipped classroom and, moreover, to be able to connect this pedagogical design with evidence of advantages related with various aspects of student learning. A review of the literature on the flipped classroom has discovered several issues, for example, (1) a large number of studies on the flipped classroom do not present their pedagogical designs with strong theoretical underpinning and are lack of pedagogical principles to guide the design, implementation and evaluation of the flipped classroom (Kim, Kim, Khera, & Getman, 2014); (2) because in-class activities vary from one to another widely, ranging from discussions, quizzes, individual/group presentations to role-play (O’Flaherty & Phillips, 2015), it is not clear what activities can help students develop critical thinking skills effectively and how; (3) due to abstract concepts or difficult contents on some learning topics, students are not able to acquire the knowledge on their own by watching lectures out of class, thus are not well prepared for in-class activities (Mason, Shuman, & Cook, 2013); (4) because of the mixed impact of the flipped learning on students’ learning outcomes and students’ perceptions of flipped learning, more research into pedagogical design is needed employing different research methods (e.g., Jungić et al., 2015).

Thus, the special issue aims to discuss pedagogical design and implementation of the flipped classroom grounded in solid theoretical principles, and evaluation of students’ learning outcomes via appropriate research methods. The ambition of this special issue is to facilitate opening the “black box” of “HOW” to design, implement and evaluate the flipped classroom, contributing insights into future directions of pedagogical practices in school and tertiary education. This special issue attracted 38 submissions with two rounds of double-blind reviews by 76 international experts. Finally, 12 papers have been selected for publication covering a wide range of topics in this field including pedagogical design, implementation and evaluation, bridging in- and out-of-class learning, learning approaches, the action research method and student perceptions of flipped learning on various subjects across school and higher education.

Designing and implementing flipped classroom in professional development remains scant. Gökçe Kurt, in her contribution entitled “Implementing the flipped classroom in teacher education: evidence from Turkey”, reports a quasi-experimental study on implementing the flipped classroom in a Classroom Management course in a pre-service English teacher education programme in higher education. The findings revealed that pre-service teachers in the flipped classroom developed more self-efficacy beliefs and performed better in their learning outcomes than those in the traditional classroom. This study provides detailed description of the pedagogical design of flipped classroom for pre-service teachers premised on the constructivist theories and differentiated instruction in its design, which enlightens future research on professional development using flipped learning approach.
Recognizing the challenges for students to comprehend the learning material on their own in flipped learning, Gwo-Jen Hwang and Chiu-Lin Lai, in their contribution entitled “Facilitating and bridging out-of-class and in-class learning: an interactive e-book-based flipped learning approach for math courses,” report a quasi-experimental study in mathematics learning in a primary school, aiming at facilitating and bridging in- and out-of-class learning using an interactive e-book-based flipped learning approach. Adopting this approach, the instructional videos, quizzes and learning guidance provided by the teacher were integrated into e-books presented on mobile devices. The results indicate that the proposed approach not only promoted the students’ self-efficacy for learning mathematics, but also improved their learning achievement, especially students with lower self-efficacy. This study contributes to the “seamless flipped learning” literature.

Although many research studies on flipped classroom have been conducted in various domain subjects, few studies have carried out in K-12 ICT (Information Communications Technology) course. Christoforos Kostaris, Stylianos Sergis, Demetrios Sampson, Michail Giannakos and Lina Pelliccione, in their contribution entitled “Investigating the potential of the flipped classroom model in K-12 ICT teaching and learning: An action research study,” present the design and implementation of an action research to examine the effect of the flipped classroom approach on K-12 ICT teaching and learning. The study provides evidence for potential advantages in students’ cognitive learning outcomes related to subject domain knowledge, the exploitation of teaching time during the classroom face-to-face sessions, the students’ level of motivation, as well as their level of engagement. It adds value to the research literature in flipped classroom regarding its first trial of the approach in the educational context of K-12 ICT teaching and integrating the four-phase of action research (Plan, Act, Observe and Reflect) into the flipped classroom.

Addressing the issue of lack of theoretical underpinning for flipped classroom pedagogical design, Chung Kwan Lo and Khe Foon Hew, in their paper entitled “Using ‘first principles of instruction’ to design secondary school Mathematics flipped classroom: the findings of two exploratory studies,” explicitly examined how flipped classroom can benefit underperforming or high ability students premised on first principles of instruction design theory in secondary school mathematics learning. Their findings show that adopting the design theory in the flipped classroom can help enhance both underperforming and high ability students' mathematics achievement. Another interesting finding is that “high ability students even asked for extending the class time in order to engage more advanced problems.” The authors, therefore, suggest that practitioners prepare extra basic exercises for underperforming students and provide high ability students more advanced and real-world problems.

In spite of comparing flipped learning approaches with the traditional teaching approaches, Morris Siu-yung Jong, in his contribution entitled “Empowering students in the process of social inquiry learning through flipping the classroom,” presents a study that integrates the idea of the flipped classroom into the process of guided social inquiry learning for promoting students’ learning achievement and self-efficacy in social and humanities education. The study shows that the proposed flipped approach has different degrees of positive effects on high, moderate, and low academic-achieving students. The approach premised on social constructivist theories adds value to the design and implementation of flipped classroom research.

Attempting a new model in flipped learning, Hsiu-Ling Chen and Chiung-Yun Chang, in their contribution entitled “Integrating the SOP2 model into the flipped classroom to foster cognitive presence and learning achievements”, report a study that explores student teachers’ cognitive presence and learning achievements by integrating the SOP2 model (“S”: Self-study, “O”: online group discussion, and “P2”: Double-stage Presentations) into the flipped classroom. The design and implementation of the SOP2 Model can be a useful reference for future adoption of the flipped classroom strategy in the higher educational context. Adopting another new approach in flipped learning, Chung-Kai Huang and Chun-Yu Lin, in their contribution entitled “Flipping business education: transformative use of team-based learning in Human Resource Management classrooms”, presents a study on integrating flipped classroom and team-based learning in two Human Resource Management classes with a two-dimensional design of pre-class, in-class and post-class activities combining with individual, team and class activities. The study shows positive relationships among students’ perceived team members’ valuable contribution, motivation, enjoyment, and learning outcomes. The findings indicate that students are more likely to learn better once they perceive that their team members are devoted to team projects.

Contrasting with the learning design of flipped classroom where direct instruction via video clips is assigned as homework task first, followed by more challenging problem solving activities in class, Yanjie Song and Manu Kapur, in their contribution entitled “How to flip the classroom – ‘productive failure’ or ‘traditional flipped classroom’ pedagogical design?” propose adopting the “productive failure-based flipped classroom” pedagogical design where students’ problem-solving goes before the instruction to enhance students’ domain knowledge and conceptual knowledge in mathematics learning in a secondary school. By inverting the traditional flipped
classroom design, this new approach is characterized by a design where students are provided opportunities to explore, discuss and solve problems related to new concepts first in class even though they might come across failures, followed by consolidating the newly learned concepts and associated procedures using video clips at home. The research results show that compared with those in the traditional flipped classroom, students in the “productive failure-based flipped classroom” made significant improvement in their conceptual understanding of solving problems. This suggests that the “productive failure” pedagogical design may be better able to improve students’ problem solving skills than traditional flipped classroom design.

Action research has hardly ever been touched on in studies of the flipped classroom. Vasiliki Aidinopoulou and Demetrios Sampson, in their contribution entitled “An action research study from implementing the flipped classroom model in primary school history teaching and learning,” report a rare case that uses flipped classroom approach for the history course adopting action research in four phases, and make this paper a very good reference for those who intend to apply flipped classroom to social studies courses using action research. They also integrated both flipped classroom and action research together and conducted this research over an entire school year with two different History classes; one followed the flipped classroom model and the other followed the traditional lecture based approach. They not only compared students’ performances on the learning goals of a history course (i.e., memorization of historical content) but also examined the effects on the cultivation of historical thinking skills (HTS). They found the engaging student-centered activities in classroom time contributed to better learning outcomes in terms of demonstrating critical HTS.

Developing students’ self-regulated learning is crucial in cultivating 21st century skills in order to succeed in the digital age. Ünal Çakiroğlu and Mücahit Özturk, in their contribution entitled “Flipped classroom with problem based activities: exploring self-regulated learning in a programming language course,” presents a study on using flipped classroom model to enhance student problem based learning and self-regulated skills in a university programming course. The study recorded student learning activities and learning skills development throughout the whole course, and demonstrated different skills development in face-to-face and home sessions. Strengths and weaknesses in applying flipped classroom model contributing to self-regulated learning for students are also discussed. This study provides insights for using problem based activities in flipped learning.

Identifying the issue of mixed research results in promoting students’ performance and/or satisfaction from flipped learning studies, Xuesong Zhai, Jibao Gu, Hefu Liu, Jyh-Chong Liang and Chin-Chung Tsai in their contribution entitled “An experiential learning perspective on students’ satisfaction model in a flipped classroom context” report on a longitudinal survey study of examining students’ satisfaction with the flipped classroom based on the experiential learning theory in higher education. The findings show that prior learning experience is a far more significant antecedent for predicting students’ satisfaction, and perceived quality (with five first-order dimensions) and perceived value are two vital mediators to students’ satisfaction.

Finally, Zandra de Araujo, Samuel Otten, and Salih Birisci, in their contribution entitled “Conceptualizing ‘homework’ in flipped mathematics classes” present their proposed framework for flipped mathematics homework. It categorizes types of homework, as well as drawing on technology literature and mathematics education literature to discern the quality for each type of homework. In addition, they show how the quality of instructional videos can vary and how teachers can assign the videos to motivate subsequent in-class work on mathematics tasks. The proposed framework provides a basis for design research focused on developing effective materials for flipped instruction in mathematics.

To conclude, “HOW” to flip the classroom? The 12 papers have addressed varied research issues from different perspectives, which provide insights into future research. We have to admit that this special issue cannot answer all the concerns related to the “HOW.” We hope that there will be more research studies to answer the question, and ultimately be able to theorize the flipped classroom to better guide the pedagogical design, implementation and evaluation of the practices.

References


Facilitating and Bridging Out-Of-Class and In-Class Learning: An Interactive E-Book-Based Flipped Learning Approach for Math Courses

Gwo-Jen Hwang* and Chiu-Lin Lai
Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan // gjhwang.academic@gmail.com // jolen761002@gmail.com
*Corresponding author

ABSTRACT

Flipped learning is a well-recognized learning mode that reverses the traditional in-class instruction arrangement by delivering learning content outside of the classroom and engaging students in more activities in class. However, it remains a challenge for students to comprehend the learning material by themselves, particularly when learning abstract concepts such as in mathematics. In this study, an interactive e-book approach is proposed to support flipped learning. It facilitates and bridges out-of-class and in-class learning by providing support for interactive learning contents presented on mobile devices. To evaluate the effectiveness of the proposed approach, a quasi-experiment was conducted in an elementary school math course. The experimental group students learned with the interactive e-book approach in the flipped learning activity, while the control group students learned with the conventional video-based flipped learning approach. The experimental results indicated that the proposed approach not only promoted the students’ self-efficacy for learning mathematics, but also improved their learning achievement; moreover, it was found that the approach benefited the lower self-efficacy more than the higher self-efficacy students. The learning record analysis further confirmed that the lower self-efficacy students spent more time reading the e-books before and in class than the higher self-efficacy students did.

Keywords
Interactive e-books, Flipped classroom, Flipped learning, Self-efficacy, Mobile technology

Introduction

In recent years, the positive effect of student-centered learning modes has been frequently discussed (Kong, 2015; McLaughlin et al., 2014). Several studies have revealed that this learning mode can improve students’ learning achievement and increase the interaction among peers and teachers (Schultz, Duffield, Rasmussen, & Wageman, 2014). Currently, flipped learning is recognized as a learning mode that achieves the goal of student-centered learning, and engages students in meaningful peer-to-peer and peer-to-teacher interactions (Gaughan, 2014; Pierce & Fox, 2012). The learning mode of flipped learning involves students watching and reviewing the learning content before taking a class. The learning content students study was traditionally taught via direct instruction, but students can actually understand the knowledge themselves. Subsequently, by adopting this approach, there is more time for students and teachers to engage in individual and small group learning (i.e., project-based learning, problem-solving, or in-depth discussion) (Gilboy, Heinerichs, & Pazzaglia, 2015).

Some of the studies have already confirmed the advantages of implementing flipped learning in regular courses. For instance, Kong (2015) conducted three years of flipped learning to improve students’ critical thinking in Humanities courses. In the flipped learning activities, the students had to preview the learning content on the pre-lesson learning platform, take part in group discussions in class, and then engage in extended learning after class. According to the intervention, it was found that the students’ critical thinking abilities were improved, and they spent more time deducting, explaining, and evaluating the knowledge related to their learning. On the other hand, Al-Zahrani (2015) integrated the flipped learning mode into an e-learning course. The findings revealed that the flipped classroom enhanced the students’ learning as well as stimulating their creativity when they were discussing or solving problems with their peers.

However, Al-Zahrani’s (2015) research also indicated the challenges of flipped learning, including the provision of adequate learning guidance. For instance, in the out-of-class learning activities, some of the learning content was abstract conceptions or was difficult to comprehend, such as mathematics concepts (Kim, Kim, Khera, & Getman, 2014; Kuo, Hwang, & Lee, 2012). Without proper guidance, students might feel helpless and fail to acquire the knowledge they need for the upcoming in-class activities (Mason, Shuman, & Cook, 2013; McLaughlin et al., 2013). Several studies have addressed this issue by analyzing students’ learning logs, and providing personalized supports via mobile devices (Ogata et al., 2015; Yin et al., 2015). Hwang, Lai and Wang (2015) further indicated the importance of bridging the out-of-class and in-class learning using mobile technology.
Therefore, in this study, an interactive e-book-based flipped learning approach is proposed. Several functions of this approach were included. For instance, the interactive e-book system consisted of all of the learning material the students have to learn outside the classroom. The students can directly make some annotations on the e-books, which they can then bring to class to share their ideas with their peers and teachers. Moreover, the system records the learning status and notes of all students. The teachers can monitor students’ learning status before starting the in-class activities. To examine the effectiveness of the proposed approach, an experiment was conducted to evaluate the effectiveness in terms of students’ learning achievement and self-efficacy. In addition, students’ learning records in the interactive e-book-based flipped learning are also analyzed and discussed.

**Literature review**

**Flipped classroom and flipped learning**

The term “flipped classroom” refers to engaging students in gaining basic knowledge before class, and providing more activities, such as doing exercises or interacting with peers and the teacher in class (Pierce & Fox, 2012). It reverses the traditional arrangement of delivering basic knowledge in class to the time outside the classroom. Moreover, it replaces the time of in-class lecturing with peer-to-peer or peer-to-teacher interactions which can effectively solve students’ learning problems and provide them with more opportunities to apply the knowledge (Bergmann & Sams, 2012; Davies, Dean, & Ball, 2013). Bergmann and Sams (2014) further defined the term “flipped learning” by emphasizing the importance of designing in-class learning activities to engage students in higher order thinking. It allows teachers to implement multiple learning strategies in their classroom, and creates a dynamic and interactive learning environment for students to apply knowledge and engage in project-based learning or inquiry learning (Bergmann & Sams, 2015).

The concept of the flipped classroom or flipped learning has been applied to various courses in recent years (Slomanson, 2014; Teo, Tan, Yan, Teo, & Yeo, 2014). For instance, Mason et al. (2013) applied the flipped classroom to a pharmaceutical course. In the study, the researchers divided the activities into two learning processes: out-of-class and in-class learning. In the out-of-class learning process, the students were asked to watch the video lecture before taking the class. In the in-class activities, the students had some group discussion and problem solving activities related to their learning contents. Mason et al. (2013) reported that the students who participated in the flipped classroom had more peer interaction in class. According to the qualitative analysis, it was found that they felt confident in learning, and experienced more engagement with the flipped classroom than with the traditional instruction. Teo et al. (2014) also conducted flipped learning in a chemistry course. There were three stages in the flipped classroom. The first was a home activity, in which the students had to watch videos and answer some related questions designed by the teacher. The second stage was the class activity. The teacher spent 15 minutes instructing the important concepts related to the students’ home activity. After that, there was laboratory work which required the students to manipulate equipment, handle materials, and make observations of the science phenomena. In the last stage, the students had to finish their laboratory report at home. This result showed that they had better understanding of the chemical theories and were able to comprehend the complex practical procedures. Moreover, the findings implied that the flipped classroom learning mode can reduce the students’ learning anxiety and improve their work efficiency (Teo et al., 2014).

However, previous studies have indicated the challenges of conducting an effective flipped classroom and flipped learning (Al-Zahrani, 2015; Gaughan, 2014). Several studies have asserted that teachers need to put a great deal of effort into providing adequate guidance for students to learn outside the classroom. With proper learning guidance, students can understand the real meaning of the learning content before class (Kim et al., 2014; Strayer, 2012). Moreover, they can concentrate on their self-directed learning process and achieve the expected goals of the flipped classroom, rather than being distracted by some useless information from irrelevant websites or platforms (Rahimi, van den Berg, & Veen, 2015).

The advancement of technologies and multimedia electronic materials (i.e., e-books, cloud computing services, handheld devices) has provided an opportunity to achieve the goals of the flipped classroom and flipped learning (Kong & Song, 2015; Hwang et al., 2015). E-books consist of various kinds of multimedia that engage students in learning with quality and well-organized content. In addition, via cloud computing services, students’ learning logs can be recorded and analyzed, and hence learning support can be provided (Hwang et al., 2015; Sandberg, Maris, & Hoogendoorn, 2014).
**E-books as a support for guiding students’ learning**

E-books represent a combination of the conventional figures in printed books and interactive computing technologies (Smeets & Bus, 2012). This kind of technology can be helpful for students who do not perform well with traditional textbooks (Maynard, 2010; Shamir & Shlafer, 2011; Yang, Hwang, Hung, & Tseng, 2013). In addition, some of the benefits of e-books have been identified: students can access digital content through e-books anytime and anywhere. E-books also offer a solution to the problem of the conventional dull instruction as they provide an opportunity for students to interact with the learning content (Jou, Tennyson, Wang, & Huang, 2016).

Previous studies have addressed several benefits of e-books integrated into education. For instance, Ihmeideh (2014) developed e-books to improve students’ emergent literacy skills. The results indicated that the students’ literacy skills were better than those of students who studied with regular printed books. In addition, the study further found that the students who learned with e-books significantly performed better in several emergent literacy aspects, such as print awareness and vocabulary. On the other hand, Huang and Liang (2015) discussed the relation of actual e-book reading behaviors and students’ comprehension outcomes. The result indicated that the reading rate of the students can reflect their reading behaviors and is further associated with students’ comprehension outcomes. This study further proved that the students’ reading process with the e-books can be analyzed and interpreted, which is much more difficult when students are studying with printed books.

According to the literature, the flipped classroom has been recognized as an innovative learning mode which leads students to apply knowledge and achieve their learning objectives (Galway, Corbett, Takaro, Tairyan, & Frank, 2014). Some research has further indicated that the incorporation of e-books in students’ learning can provide adequate guidance and improve their learning performance (Strayer, 2012). In addition, the students’ learning behaviors can be easily recorded, which assists teachers and educators in understanding their learning in depth.

Therefore, in this study, an interactive e-book-based flipped learning approach is proposed for assisting students’ learning outside of the classroom, and helping them bring their knowledge to their in-class discussion. Moreover, some research questions are investigated to evaluate the effectiveness of the proposed approach:

- Can the interactive e-book-based flipped learning approach improve the students’ learning achievement in comparison with the conventional video-based flipped learning?
- Can the interactive e-book-based flipped learning approach improve the students’ self-efficacy in comparison with the conventional video-based flipped learning?
- Are there differences between the learning frequency of the lower self-efficacy and higher self-efficacy students in learning with the interactive e-book-based flipped learning approach?

**Interactive e-book environment for flipped learning**

In this study, an interactive e-book-based flipped learning environment was developed with HamaStar SimMAGIC Designer and HamaStar SimMAGIC CLibrary. The structure of the proposed approach is composed of three components, namely cloud equipment, cloud resources, and cloud services, as shown in Figure 1. The cloud equipment consists of the system software, physical hardware, and network devices and environment. More specifically, the system software refers to the software which the teachers can use to develop their courses related to e-books, while the platform refers to where they can manage their online learning contents. Besides, the cloud resources provide the students with the learning material, the e-books, and related learning tasks. Those learning materials and resources on the cloud resources layer can be uploaded and managed by the teachers. Finally, the cloud services provide public or personal information to both teachers and students. The public information consists of the e-books that are available for every user, while the personal information represents the individuals’ readings, their personal learning logs, and performance of the tasks. The information sharing service is also provided for the students or teachers to share their notes or tasks with a specific user. In the cloud services layer, the students can manage their readings and see their learning logs. In the meantime, the teachers can also monitor their students’ learning logs and assign adaptive readings for a particular student.

Figure 2 shows the interface of interactive e-books. When the students log into the e-book system, they can search for the books they want to read or check the “assigned books” to access the e-books they are required to read. Once the e-books have been downloaded to the students’ personal tablets, they can read them offline in “my bookcase.”
Figure 1. Structure of the e-book library supporting the flipped learning

Figure 2. Interface of the e-book library

Figure 3. Interface of the e-books
As soon as the students select an e-book to read, it opens, as shown in Figure 3. A toolbar is hidden on the left side of the e-book, so the students can open it when they need it. A number of functions are included in the toolbar, such as bookmark, search, brush, text, take notes, share note, etc.

Figure 4. The e-book knowledge description and video watching functions

Figure 5. The interactive feedback provided to the students

Figure 6. Interface of sharing notes during the learning process
The students can not only read the learning content and complete the learning tasks, but can also take some notes during the learning process. For instance, they can use the brush to mark any concepts they have trouble understanding, and they can write some notes on the e-books. In the meantime, they can share their notes with their peers. More importantly, they are allowed to bring the e-books along with the notes they have taken to class so as to facilitate their learning in the forthcoming in-class activities. Figure 6 shows the interface which allows the students to read others’ notes. The students can choose the function “read others’ note” on the toolbar and select a specific student’s notes. The system then downloads the selected student’s e-book including his/her notes, so that the students can directly read the e-book and the notes made by others.

The e-book system also provides an additional function for teachers to realize the learning status of the students, so that proper in-class remedial instruction or activities can be provided accordingly. For example, the number of mistakes students make in solving each math question at home after learning with the e-book are recorded and summarized to help teachers determine whether additional instructions are needed in the class, as shown in Figure 7.

![Figure 7. Summary of students’ at-home learning status](image)

**Experimental design**

To evaluate the effectiveness of the interactive e-book-based flipped learning approach, an experiment was conducted in an elementary school math course to investigate the effects of the approach on the students’ learning achievement and self-efficacy. Moreover, the students’ learning records were analyzed as well.

**Participants**

The participants of this study included two classes of fourth-grade students. One class was assigned to the experimental group and the other was the control group. The experimental group consisted of 24 students learning with the interactive e-book-based flipped learning; that is, the instructional videos, quizzes and learning guidance provided by the teacher were integrated into e-books. On the other hand, the control group students (n = 21) learned with the conventional video-based flipped learning approach. They learned with the instructional videos with printed learning sheets provided by the same teacher.
Instrument

The measurement instruments adopted in this study included the pre-test, post-test, and the questionnaire of self-efficacy.

The pre-test and post-test were developed by three experienced teachers. The pre-test aimed to evaluate whether the two groups of students had equivalent prior knowledge of the “Area and perimeter” concept of their math courses. It consisted of 5 multiple-choice items, 2 matching items, 7 fill-in-the-blank items, and 10 question-and-answer items, with a perfect score of 100. The post-test aimed to evaluate the students’ knowledge and competence of identifying various volumes and calculating complex volumes. It consisted of 10 multiple-choice items (50%), 5 matching items (25%), and 5 question-and-answer items (25%), and the perfect score of the post-test was 100.

On the other hand, the questionnaire of self-efficacy was modified from the measurement developed by Wang and Hwang (2012). It consisted of eight items with a five-point Likert scale. This questionnaire can be utilized for measuring the students’ expectations and confidence regarding learning the math course content well (e.g., “I believe that I can understand the most difficult part of this course”). The Cronbach’s alpha value of the questionnaire was 0.92, implying that the questionnaire is reliable.

Experimental procedure

Figure 8 shows the four-week experimental procedure. In the first week, the students in the two groups took the pre-test and completed the self-efficacy pre-questionnaire. Moreover, the teacher introduced the syllabus and the learning objectives. After that, the experimental group students learned with the interactive e-book in their out-of-class time, while the control group students learned with the conventional online videos with learning sheets provided by the teacher.

Following the school curriculum arrangement, there were two 50-minute periods of in-class learning activities each week. In the in-class activities, the experimental group students brought their tablet computers with the annotated e-books, while the control group students brought their printed learning sheets with the notes they had taken. Both groups of students participated in the group discussion based on the same guideline provided by the teacher. In addition, they were asked to raise problems they had encountered in the out-of-class activities. Then, the teacher would encourage other classmates to solve the problems using electronic whiteboards, as shown in Figure 9. In the fourth week, the students took the post-test and the post-questionnaire of self-efficacy.

Figure 8. Diagram of the experiment design
Results

In this study, an interactive e-book-based flipped learning approach was proposed to support students in a flipped learning context. In order to evaluate the effects of this approach, a quasi-experiment was conducted to investigate the effects of the proposed approach on students’ learning achievement and self-efficacy. Moreover, the students’ learning behavior regarding the e-books is further discussed.

Analysis of learning achievement

In order to examine the effects of the proposed approach, the students’ learning achievement before and after the experiment was explored. The mean values and standard deviations of the pre-test scores were 82.69 and 9.67 for the experimental group students, and 80.57 and 17.79 for the control group students. The t-test shows that there was no significant difference between the two groups (t = 0.50, p > .05), indicating the equivalent prior knowledge of the two groups of students before engaging in the learning activity.

After the students finished the flipped learning activities, the post-test was conducted to evaluate the students’ learning knowledge and capabilities of identifying various volumes and calculating those volumes. Table 1 shows the t-test result of the post-test scores of the two groups. The means and standard deviations were 80.92 and 5.69 for the experimental group students, and 74.76 and 10.31 for the control group students. The t-test result indicated that the post-test scores of the two groups were significantly different (t = 2.43, p < .05, Cohen’s d = 0.74). Furthermore, the effect size (d) was 0.74, showing a medium to large effect (Cohen, 1988).

Table 1. The t-test result of the post-test scores of the two groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>T</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-test</td>
<td>Experimental</td>
<td>24</td>
<td>80.92</td>
<td>5.69</td>
<td>2.43*</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>21</td>
<td>74.76</td>
<td>10.31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05.

On the other hand, the students’ self-efficacy can be regarded as the criteria of their capabilities of organizing and executing the course and further achieving certain goals of learning, which has high correlation with their learning achievement (Bandura, 1977; Wang & Wu, 2008). Therefore, this study further explored the effects of this approach on the learning achievement of the students with different self-efficacy levels. A two-way ANOVA was further employed to evaluate the interaction between the different learning approaches (interactive e-book-based flipped learning and conventional video-based flipped learning) and different self-efficacy levels (higher and lower). The students were classified into high and low self-efficacy groups based on their self-efficacy ratings on the pre-questionnaire. The students in the top 50% were regarded as having higher self-efficacy, while the others had lower self-efficacy. The independent variables were the two learning approaches (i.e., interactive e-book-based flipped learning and conventional video-based flipped learning) and two levels (i.e., higher and lower) of self-efficacy, while the dependent variable was the students’ learning achievement. The assumption of homogeneity of regression was not violated (F = 2.25, p > .05), suggesting that a common regression coefficient was appropriate for the two-way ANOVA.
Table 2 shows the two-way ANOVA results of the learning achievement. It was found that significant effects were observed for the learning approaches ($F = 5.27, p < .05$), self-efficacy ($F = 6.08, p < .05$), and the interaction between the learning approaches and self-efficacy ($F = 4.81, p < .05$) on the students’ learning achievement. Furthermore, the effect sizes ($\eta^2$) were 0.11, 0.13, and 0.11 for each variable, representing a moderate effect size (Cohen, 1988).

<table>
<thead>
<tr>
<th>Variable</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning approach</td>
<td>292.49</td>
<td>1</td>
<td>292.49</td>
<td>5.27</td>
<td>0.03</td>
<td>0.11</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>337.12</td>
<td>1</td>
<td>337.12</td>
<td>6.08</td>
<td>0.02</td>
<td>0.13</td>
</tr>
<tr>
<td>Learning approach*Self-efficacy</td>
<td>266.93</td>
<td>1</td>
<td>266.93</td>
<td>4.81</td>
<td>0.03</td>
<td>0.11</td>
</tr>
<tr>
<td>Error</td>
<td>2274.31</td>
<td>41</td>
<td>55.47</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A simple main-effect analysis was further employed to explore the effects of the self-efficacy levels on the learning achievement of the students who learned with different flipped learning approaches, as shown in Table 3. It was found that the students with different self-efficacy levels in the control group showed significantly different learning achievement ($F = 10.66, p < .01, \eta^2 = 0.21$). On the other hand, there was no significant difference between the students in the experimental group who had different levels of self-efficacy ($F = 0.94, p > .05$). These results indicate that the engagement of students’ self-efficacy levels only played an important role when students were learning with the conventional video-based flipped learning. In other words, with the assistance of interactive e-books, fewer effects of the students’ learning efficacy on their achievement would occur.

<table>
<thead>
<tr>
<th>Variable</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive e-book-based flipped learning (experimental group)</td>
<td>Between groups</td>
<td>2.08</td>
<td>1</td>
<td>2.08</td>
<td>0.04</td>
<td>0.85</td>
</tr>
<tr>
<td>Conventional video-based flipped learning (control group)</td>
<td>Between groups</td>
<td>591.25</td>
<td>1</td>
<td>591.25</td>
<td>10.66</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Table 4 shows the simple main-effect analysis results of the effects of the learning approaches on the learning achievement of the students with different self-efficacy levels. It was found that the lower self-efficacy students showed significantly different learning achievements when learning with different approaches ($F = 9.24, p < .01, \eta^2 = 0.18$), while there was no significant difference between higher self-efficacy students with different learning approaches ($F = 0.01, p > .05$). According to the results, it is implied that the interactive e-book-based flipped learning approach could benefit the students who engaged in lower self-efficacy more than those with higher self-efficacy.

<table>
<thead>
<tr>
<th>Variable</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher self-efficacy</td>
<td>0.32</td>
<td>1</td>
<td>0.32</td>
<td>0.01</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>Within groups</td>
<td>2274.31</td>
<td>41</td>
<td>55.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2274.63</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower self-efficacy</td>
<td>512.53</td>
<td>1</td>
<td>512.53</td>
<td>9.24</td>
<td>0.004</td>
<td>0.18</td>
</tr>
<tr>
<td>Within groups</td>
<td>2274.31</td>
<td>41</td>
<td>55.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2786.84</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 10 shows the interaction between the different flipped learning approaches and the self-efficacy levels on the students’ learning achievements. It was clearly found that the students who learned with the interactive e-book-based flipped learning performed significantly better than those who learned with the conventional flipped classroom approach. Moreover, while learning with the conventional video-based flipped learning, the students who had higher self-efficacy showed significantly higher achievement than those with lower self-efficacy. These results imply that students’ self-efficacy plays an important role when they are learning in the conventional video-based flipped learning context. Their self-efficacy performance has a certain positive impact on their learning achievement. On the other hand, there was no difference found in the students’ learning achievement in the different self-efficacy levels in the experimental group, indicating that the interactive e-book-based flipped learning approach can support students’ learning and improve their learning achievement, regardless of whether the students have higher- or lower-level self-efficacy.
Analysis of self-efficacy

In order to evaluate the effects of the interactive e-book-based flipped learning on students’ self-efficacy beliefs, in this study, the students’ self-efficacy before and after the experiment was collected. Furthermore, a one-way ANCOVA examination of the students’ self-efficacy in the experimental group and the control group was adopted. The assumption of homogeneity of regression was not violated \( F = 2.11, p > .05 \), showing a common regression coefficient for one-way ANCOVA. Table 5 shows the results of self-efficacy for the two groups. The adjusted means of the values were 3.98 in the experimental group and 3.60 in the control group. The standard errors of the two groups were 0.12 and 0.13, respectively. According to the ANCOVA result, it was found that the self-efficacy of the experimental group was significantly higher than that of the control group \( F = 4.57, p < .05, \eta^2 = 0.10 \). This result implies that the interactive e-book-based flipped learning approach can stimulate the students to have stronger beliefs in learning than the conventional flipped classroom.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted Mean</th>
<th>Adjusted SD</th>
<th>F</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>24</td>
<td>4.00</td>
<td>0.66</td>
<td>3.98</td>
<td>0.12</td>
<td>4.57</td>
<td>0.10</td>
</tr>
<tr>
<td>Control group</td>
<td>21</td>
<td>3.58</td>
<td>0.51</td>
<td>3.60</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* \( p < .05 \).

Analysis of the e-book learning records

In order to understand the students’ learning behaviors in the interactive e-book-based flipped learning, this study further analyzed the experimental group students’ learning records in the interactive e-books. The interactive e-book system automatically recorded the students’ reading frequency and reading time of each interactive e-book. This study, therefore, classified the students’ learning records based on their ratings of self-efficacy on the pre-questionnaire (higher self-efficacy and lower self-efficacy). The values of the students’ self-efficacy higher than the mean value (3.81) were regarded as higher self-efficacy, while the others were considered as lower self-efficacy. Moreover, we separately discuss the students’ learning frequency in the first period of the flipped classroom (unit 1 e-book reading) and the second period (unit 2 e-book reading). Figure 11 shows the learning records of the students in the different self-efficacy levels for each e-book, where Si represents student identification.

According to the learning records, the students with higher self-efficacy spent 19.57 minutes on average reading the unit 1 e-book, while the students with lower self-efficacy spent 18.05 minutes on average. On the other hand, 72% of the higher self-efficacy students read the unit 1 e-book more than two times, while 75% of the lower self-efficacy students read the unit 1 e-book more than two times.

Figure 11. Interaction between self-efficacy levels and learning approaches
Compared to the unit 1 reading, the students with higher self-efficacy spent 17.80 minutes on average reading unit 2. Just over half of the higher self-efficacy students (55.56%) read the unit 2 e-book more than two times, which was significantly less than their reading of the unit 1 e-book. The lower self-efficacy students, on the other hand, spent 21.05 minutes on average reading the unit 2 e-book, and more than 62.50% of the lower self-efficacy students read the unit 2 e-book more than two times.

According to the learning records, some conclusions can be drawn. First, no matter whether the students had higher or lower self-efficacy, far fewer students read the unit 2 e-book more than two times compared with the unit 1 e-book. Second, the average reading time of the two e-books slightly decreased for the higher self-efficacy students, while it slightly increased for the lower self-efficacy students. Third, the proportion of the lower self-efficacy students who read the unit 2 e-book more than two times was higher than the proportion of higher self-efficacy students who did.

**Discussion and conclusions**

In recent years, the implementation of flipped learning has been frequently discussed. Some of the research has proved the advantages of flipping the classroom, and its benefits for the students' learning performance (Hung, 2015; Wanner & Palmer, 2015). However, several studies have indicated the importance of providing multimedia and proper learning guidance for supporting and bridging the in-class and out-of-class activities (Al-Zahrani, 2015; Strayer, 2012). In order to achieve this purpose, in this study, an interactive e-book-based flipped learning approach was proposed for students to learn abstract mathematical conceptions. In this learning mode, students can read multimedia materials and preview the learning content in the interactive e-book system, and make their own annotations on the e-books. In the in-class activities, the students can bring all the learning content and annotations they have made to class, and have in-depth discussions with their teachers or classmates. An experiment was conducted in an elementary school to examine the proposed learning approach. The experimental group used the interactive e-book-based flipped learning approach, while the control group learned with the conventional video-based flipped learning approach. The experimental results showed that the proposed approach significantly benefited the students’ learning achievement and self-efficacy.

This finding proved that the interactive e-book system can benefit students in a flipped learning context, and encourages students to construct knowledge by themselves. This result is consistent with previous studies which found that the provision of e-books in learning can improve students’ learning achievement and increase their confidence in learning (Huang, Liang, Su, & Chen, 2012). Moreover, the findings of this study suggest the need to provide proper guidance and support for bridging out-of-class and in-class learning (Hung, 2015; McLaughlin et al., 2013).

According to the results of simple main-effect of self-efficacy levels on students’ learning achievement, it was found that the interactive e-book-based flipped learning improved the students’ learning achievement regardless of whether they had higher- or lower-level self-efficacy. In the conventional video-based flipped learning, it was found that the higher self-efficacy students outperformed the lower self-efficacy students. These results indicate...
that students’ self-efficacy in math learning plays an important role in flipped learning, and teachers need to pay more attention to those students with lower self-efficacy when conducting conventional flipped learning.

On the other hand, the simple main-effect of learning approach on students’ learning achievement revealed a significant difference for lower self-efficacy students who learned with the different learning approaches, but this difference was not found for the higher self-efficacy students. In the social cognitive theory, Bandura (1986) described the reciprocity of learners’ behaviors, personal factors and environments, and indicated that environmental factors would determine students’ learning behaviors and influence their efficacy beliefs (van Dinther, Dochy, & Segers, 2011). In the current study, the environmental intervention of the interactive e-books approach also confirmed this theory. Since the students who had higher self-efficacy already had higher cognitive engagement in learning, it was certain that the higher self-efficacy students were able to adopt better strategies for learning (Tsai, Ho, Liang, & Lin, 2011). On the contrary, students who have lower self-efficacy usually engage in self-devaluation, and are depressed or may withdraw from learning (Bandura, 2009). However, in this study, it was interesting to find that the learning achievement of the lower self-efficacy students was as good as that of the higher self-efficacy students when learning with the interactive e-book-based flipped learning. The e-book logs further showed that the lower self-efficacy students’ learning frequency for unit 1 and unit 2 was stable. Moreover, they spent more time learning unit 2. These results reveal that the interactive e-book-based flipped learning approach stimulated the lower self-efficacy students to spend more time reading.

The effectiveness of the proposed approach could be due to the fact that it successfully served as guidance to support and bridge the students’ learning outside and within the classroom; this was indicated by Hwang et al. (2015) who argued that “seamless flipped learning” could benefit students by improving their learning performance. In the current study, the annotation function and cloud services in the interactive e-book environment allowed the students to bring all of the learning content, the annotations they made, and the questions they raised during the out-of-class and in-class activities with them. Therefore, they could seamlessly and continuously review, construct, and practice their learning knowledge across the learning contexts.

In sum, the main contribution of this study is to show the effectiveness of “seamless flipped learning,” which refers to the use of mobile technology to bridge the out-of-class and in-class learning by providing learning supports in the class based on the students’ out-of-class learning status, as well as allowing them to bring what they have learned at home to the in-class activities. In the future, note-recommendation and advanced visualization functions can be implemented in the e-book system based on learning log analysis to better assist students and teachers, as indicated by several researchers (Aljohani & Davis, 2012; Mouri & Ogata, 2015).

Acknowledgements

This study is supported in part by the Ministry of Science and Technology of Taiwan, R.O.C. under contract numbers MOST 104-2511-S-011-001-MY2 and MOST-105-2511-S-011 -008 -MY3.

References

Aljohani, N. R., & Davis, H. C. (2012). Significance of learning analytics in enhancing the mobile and pervasive learning environments. In Sixth International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST) (pp. 70-74). doi:10.1109/NGMAST.2012.49


Gaughan, J. E. (2014). The Flipped classroom in world history. *The History Teacher, 47*(2), 221-244.


An Experiential Learning Perspective on Students’ Satisfaction Model in a Flipped Classroom Context

Xuesong Zhai1,3, Jibao Gu1,*, Hefu Liu1, Jyh-Chong Liang2 and Chin-Chung Tsai2
1School of Public Affairs, University of Science and Technology of China, Hefei, China // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan // 3School of Foreign Studies, Anhui Jianzhu University, Hefei, China // zhxs@mail.ustc.edu.cn // Jibao@ustc.edu.cn // liufh@ustc.edu.cn // aljc@mail.ntust.edu.tw // cctsai@mail.ntust.edu.tw
*Corresponding author

ABSTRACT
Recent years have witnessed an increasing interest in the flipped classroom model, and many flipped programs have been funded and implemented to explore the effectiveness of this new model. However, previous studies centering on comparative assessment have indicated that it is not always entirely successful in terms of promoting students’ performance and/or satisfaction, which warrants further research on the contributing factors and driving mechanism accounting for students’ perceptions of flipped settings. In order to fill this gap, in this study, a students’ satisfaction model for the flipped classroom was constructed based on the experiential learning theory. A total of 178 undergraduate students in Mainland China participated in 32-week College English flipped classes, from whom 146 valid questionnaires were obtained. The proposed research model was evaluated through longitudinal surveys followed by the structural equation modeling technique. The results indicated that, compared with the designs of Personalized Learning Climate, learners’ Prior Learning Experience is a far more significant antecedent for predicting their satisfaction. Furthermore, Perceived Quality (with five first-order dimensions) and Perceived Value are two vital mediators to student satisfaction. The implications of this study are also discussed.

Keywords
Flipped classroom model, Satisfaction, Experiential learning, Prior learning experience, Personalized learning

Introduction
Currently, the Flipped Classroom Model (FCM), featuring especial emphasis on students’ engagement and experience (Bergmann, Overmyer, & Wilie, 2011), is increasingly attracting educators’ interest, resulting in the flipped classroom phenomenon (Blair, Maharaj, & Primus, 2015). Many colleges and universities are funding and developing FCM programs focused on comparative assessments of students’ examination scores and/or attitudes (Tune, Sturek, & Basile, 2013; Schultz, Duffield, Rasmussen, & Wageman, 2014; Baepler, Walker, & Driessen, 2014; Kong, 2014). However, the results vary greatly, which aroused our curiosity about what factors drive effective flipped pedagogy, and how they can be efficiently implemented in further teaching practice.

The learners’ satisfaction model warrants keen interest in this new and exciting research field, because learners’ perceived satisfaction has been proven to be a vital predictor of learning outcomes and behavioral intention to continue learning (Tsai, Lin, & Tsai, 2001; Liaw, 2008). In addition, student satisfaction has a close relationship with learners’ active participation and team collaborative learning (Johnson, Top, & Yukselturk, 2011; Ku, Tseng, & Akarasriworn, 2013), which is of great significance in effectively implementing the flipped classroom pedagogy. However, the existing studies ignore the fact that learners’ satisfaction should be especially discussed through the lens of learners’ experiential learning, the main feature of FCM. Some researchers have also urged further studies on the activity-oriented nature of flipped learning when exploring the learner satisfaction model in FCM (Chen, Wang, & Chen, 2014).

Experiential Learning Theory (ELT) is an applicable theoretical foundation to investigate learners’ satisfaction in flipped settings. Based on the ELT, we proposed that personalized learning climate (flipped design) and relevant prior learning experience (angle from learners) have close relationship with learner satisfaction. For one thing, there exist no one-fits-all approach for students who have distinct learning capacities and styles; thus, experiential learning especially addresses the importance of creating a personalized learning climate to meet individuals’ special needs (Sims, 2002), and the flipped settings offering learners more flexible learning arrangements are expected to relate to learners’ satisfaction. Secondly, learners’ relevant prior learning experience may significantly predict their satisfaction in flipped settings. The ELT proposed that learners’ relevant prior learning experience, such as information retrieval and online interaction, are available inputs for improving learners’ personal or group effectiveness (Kohonen, Jatainen, Kaikkonen, & Lehtovaara, 2014). Based on ELT, some researchers have also appealed for more opportunities for students to implement reflection and reconstruction of
previous experience and new ideas (Oxley & Ilea, 2015), which is in accordance with the philosophy of flipped pedagogy.

In order to further explore the mechanism of how these two factors predict students’ satisfaction, the current research employed perceived quality and perceived value as two mediators in the proposed model. The former addresses the assessment of the course content, while the latter centers on learning efficiency which sheds light on whether the course contents are effectively organized and implemented. In a flipped context, personalized learning procedures and a wealth of blended learning experience are helpful for learners to comprehend the course contents and to master the techniques of learning, which is followed by improving learners’ satisfaction. It has also been suggested by previous studies that perceived quality and perceived value act as significant mediators of learner satisfaction (Shi, 2010; Lee, 2010).

The current survey empirically aims to construct a theoretical model which is reliable for predicting undergraduates’ satisfaction in FCM context from the perspective of learners’ experiential learning. Additionally, for the practical implication, the current research can assist educators in effectively utilizing and improving the flipped approach to enhance learning perceptions and outcomes.

**Research background and hypotheses**

**Flipped classroom model**

Distinct from the traditional lecture-plus-homework formula, in the FCM context, the externalization of knowledge is previously carried out through online courses, information searching and online discussion at the learners’ own pace, rather than passively receiving information. On the other hand, the internalization of knowledge is conducted in face-to-face classes through collaborative work (such as discussion, argument and retrospection) among peers and instructors to solve targeted problems. Besides, online interactive platforms (e.g., chat rooms, BBSs) are used for the online exchange of ideas (Marcey & Brint, 2012; Chen, Wang & Chen, 2014). Synthesizing the above definition, we proposed that FCM consists of three basic pillars: (1) online videos, (2) physical classes, and (3) an interactive platform. Compared with the traditional blended learning model, the FCM replaces the passive, didactic format with a model of initiative and personalized education. It creates a student-centered learning atmosphere involving problem-based learning (PBL) modules to enhance critical thinking and self-directed learning skills (Mason, Shuman, & Cook, 2013).

Prior literature concerning FCM learners’ satisfaction has indicated that this approach is not always completely successful in promoting students’ favorable perceptions. Some surveys have found that the students’ perceptions were far more positive towards the FCM than the traditional approach (Butt, 2012; Marlowe, 2012; Baepler, Walker & Driessen, 2014). Nevertheless, Blair, Maharaj, & Primus’s (2015) survey found no significant change in relation to students’ attitudes, while Bishop and Verleger (2013) even found that a few students strongly disliked the flipped model.

Some prior research was dedicated for exploring the factors driving learners’ varied perceptions. Bergmann and Sams (2012) emphasized the importance of content delivery in their Flipped Mastery Model, while Davies, Dean, and Ball (2013) found that the effective application of technology in a flipped classroom is a vital predictor of learners’ motivation. Another survey indicated that active learning classrooms with efficient use of physical space have close relationships with learners’ perceptions of flipped settings (Baepler, Walker & Driessen, 2014). However, prior studies have been mainly confined to the perspective of course components. The exploration of learners’ satisfaction must be from the perspective of students’ experiential learning, as a result of which FCM especially features stress on individual arrangement and students’ engagement. There is thus an urgent need to construct a holistic satisfaction model from the perspective of the students’ learning experience, with which future flipped pedagogy practice could be improved.

**Theory and hypotheses**

*Experiential learning theory*

Experiential learning theory, developed from the work of John Dewey and Kurt Lewin, is applicable to the theoretical foundation of this study. The core concept of ELT emphasizes that learners’ experience has great potential to contribute to knowledge construction and comprehension (Kolb, 1984). Experiential learning,
involving a creative tension among four learning modes—concrete experience, reflective observation, abstract hypotheses, and active testing, follows a repeated cycle of continuous experience and exploration, and is portrayed as a spiral learning cycle during which the learner “touches all the bases” (Kolb & Kolb, 2005).

The procedure of FCM is in accordance with the ideology of experiential learning. Learners firstly consolidate their experiences through online courses, at which stage there possibly exist some difficulties or speculations on the part of the learners. Then online chat and a Q&A platform are utilized for reflective observation, during which the pre-viewed online lectures could be better understood. The learners’ perception of quality is thus improved. However, individuals’ understanding and thinking may not reach the consensus, and the abstract conceptualization of their comprehension requires sharing and mutual consultation among instructors and peers. At this stage, the learners not only get their personalized solution but also learn more from the instructor’s and peers’ diverse answers. Their perceived value may then be enhanced. When reflections are summarized into abstract concepts under the instructor’s guidance and peers’ mutual feedback in physical classes, new implications can be drawn. This is then followed by testing what they have obtained. Finally, they prepare to engage in the next experiences.

Prior learning experience and student satisfaction

Experiential Learning Theory emphasizes the significance of learners prior experience – such as technology utilization, teaching style – to the effectiveness of their study, as learners’ prior experience would generate their reflection, which could then be applied to new contexts to guide their learning activities (Chen, Wei, Liu, 2011).

Additionally, the ELT considers learning as a process carried out under the stimuli of the learner’s own direct experience or reacting from external observation, and knowledge is created through the transformation of experience. Then prior experience not only simply directs a person, but also affects the formation of their attitudes, desire and purpose (Kolb, 2014). Xu et al. (2014) also posited that the educational situation is supposed to be modified according to the learners’ prior experience. In the flipped context, learning effectiveness and preferences largely rely on learners’ previous technical mastery, such as the application of online courses and interactive platforms, which is in accordance with the implications of prior studies (Sahin & Shelley, 2008; Teo, 2013). Moreover, prior learning experience in the current research ought to be defined with new conceptions, considering that the procedure of flipped classes is integrated by on-and-off learning models, or blended learning, from which learners’ participation frequency and preference for blended learning constitute their prior learning experience in flipped settings. Synthesizing these thoughts leads to the following hypothesis:

**H1.** Prior learning experience is positively associated with students’ satisfaction in FCM.

Personalized learning climate and student satisfaction

Personalized learning climate could be perceived as a pedagogical approach for adjusting to students’ learning style and pace (Dabbagh & Kitsantas, 2012), which empowers students to select favored techniques and materials to efficiently manage, create and package learning content (McLoughlin & Lee, 2010). The flipped procedure provides the learners with an environment in which the learning arrangement could be adjusted to their own learning pace, approach and interests. Specifically, before the class, students can operate the video lectures (forward, replay, or pause to search for information) by adjusting themselves to their preference, and they do not need to worry about their inconsistent learning progress with other classmates. Such a fair atmosphere helps to promote their psychological wellbeing; students also feel more confident in the physical classroom as a result of the fact that their target problems have been previously prepared and negotiated, and the online platform is readily available for their instant recourse. Thus, it is suggested that a Personalized Learning Climate (PLC), the typical feature of flipped settings, is a significant predictor of learners’ perceptions. Earlier research has supported that the correlation between learners’ satisfaction and the personalized system is very high (Chen, Lee, & Chen, 2005). An empirical field experiment conducted by Xu, Huang, Wang, and Heales (2014) found that personalized e-learning facilities can enhance online learning effectiveness in terms of examination, satisfaction, and self-efficacy, since personalized settings offer learners an environment in which their ideas can be explored, compared and critiqued. Synthesizing these findings leads to the following hypothesis:

**H2.** Personalized learning climate is positively associated with students’ satisfaction in a flipped classroom.
**Perceived quality in FCM and its role as mediator**

Based on the definition of FCM, it consists of three basic pillars – online video, face-to-face discussion, and interaction with the online platform – all of which should be considered to contribute to the learners’ perceived quality. Previous research has concluded that in e-learning contexts the effect of perceived quality is mainly measured by ease of using the information system, the reliability of using information technology, and interactive response (Chiu et al., 2005). While in the settings of flipped classrooms the dimensions of interactive platforms and online courses have a close relationship with information technology, additionally flipped settings emphasize effective interaction among instructors and peers in physical classes for promoting teaching quality. Thereby, students’ Perceived Quality in the current research is proposed to be constructed of five assessment factors: (1) Ease of using the online course (EUOC); (2) Usefulness of the online course (UOC); (3) Ease of using the platform (EUP); (4) Usefulness of the platform (UP); and (5) Interaction in physical classes (IPC).

There is a substantial amount of literature elaborating that perceived quality is a vital driver of learners’ satisfaction (Helgesen & Nesret, 2007; Ribbink, Van Riel, Liljander, & Streukens, 2004). Such a mechanism, we suppose, is still available in flipped pedagogy, since the perceptions of ease of use and usefulness of information technology are generated through long-term learning experiences, and learners have such a deep understanding of them that their overall perceptions can be justified.

Perceived Quality is proposed to mediate Personalized Learning Climate/Prior Learning Experience and Student Satisfaction in flipped settings. Firstly, considering that FCM is such a complex integration of educational technology and flipped settings that learners certainly have a vague concept of it, learners’ prior experience concerning blended learning helps to improve their perceptions of the ease of use and usefulness of the interactive platform and online course. Besides, students with more experience of internet-based courses have been found to perceive higher capability and appeared to be more interested in collaborative work (Lee & Tsai, 2011). Prior blended learning experience is therefore considered to highly predict Perceived Quality. Secondly, in flipped settings, learners can speed up/slow down the learning pace, adjusting how much they learn according to their own learning progress, which is helpful for them to better understand the lessons, and more easily submit, search for, and collect needed information. Similarly, learners in physical classes would perceive more effective interactions when their targeted personal problems are settled. Therefore, the following hypotheses are proposed:

- **H3a.** Perceived quality is positively associated with student satisfaction in FCM.
- **H3b.** Prior learning experience is positively associated with perceived quality in FCM.
- **H3c.** Personalized learning climate is positively associated with perceived quality in FCM.

**Perceived value and its role as mediator**

Perceived Value, defined as perceived assessment of “how much received versus cost,” is a vital measurement indicating learning efficiency, which needs to be carefully considered when evaluating learners’ perceptions (Chiu et al., 2005). Meanwhile, distinct from general commodities, educational activity deserves to be assessed by consumed time rather than charged expense. The ratings of consumed time and learning effectiveness were thus utilized to assess learners’ perceived value in this study.

Learners’ perceived value is positively related to their satisfaction (Chiu et al., 2005). When determining to adopt a new learning model, such as flipped pedagogy, students are likely to expect to achieve their study mission more efficiently, and they may be sensitive to how much they receive versus how much they spend. Once such an expectation is satisfactorily achieved, their overall satisfaction would then be greatly improved. Additionally, relevant previous learning experience may also contribute greatly to the enhancement of perceived value. In flipped settings, individuals unfamiliar with the learning settings inevitably need time to cultivate the study techniques required to exploit the new technology (hardware and software) and blended learning skills. Users more familiar with them are more likely to identify the requisite learning skills and to become proficient in their use (Owens, Hardcastle, & Richardson, 2009). Besides, earlier research has indicated that students learning with a personalized learning climate had higher learning effectiveness (Mohd, Shahbodin, Pee, & Hanapi, 2013). In flipped pedagogy, personalized assistance supported by an interactive platform is beneficial for learners to meet their specific needs timely, and their perceived value will be further enhanced as their target individual issues are dealt with in physical classes. Based on the above analysis, we propose that:

- **H4a.** Perceived value is positively associated with student satisfaction in FCM.
- **H4b.** Prior learning experience is positively associated with the perceived value of FCM.
- **H4c.** Personalized learning climate is positively associated with perceived value in FCM.
Perceived quality and perceived value

We suppose that perceived quality in the flipped contexts has a positive effect on perceived value. Specifically, since perceived quality is measured by five first-order factors in the current research, perceived usefulness of the online course relates to the perception of the course content, and the dimension of ease of using the online course makes individuals feel that their time management is economical, both of which together have an influence on their perceptions of the learning efficiency of the online course. Prior research also supports that perceived ease of use and usefulness would positively relate to students’ learning efficiency in blended settings (Teo, 2010). This deduction also applies to the learners’ perceptions of using the online information platform. Moreover, in physical classes, collaborative learning among teachers and peers can efficiently promote students’ achievement and co-construction of knowledge (Tsai, 2001). We thus propose that:

H5. Perceived quality is positively associated with perceived value in FCM.

Proposed FCM satisfaction model

Synthetically considering the above hypotheses, the experiential learning theory and the features of the flipped classroom, this research proposed personalized learning climate and prior learning experience as significant antecedents to predict learners’ satisfaction; furthermore, perceived quality and perceived value were introduced to further explore their vital mediation mechanism in the proposed model (shown in Figure 1).

Methodology

Participants and procedure

A total of 178 undergraduates participated in the Flipped English as Foreign Language (EFL) sample class at a target University in central Mainland China. The longitudinal survey methodology was used for collecting data. The measurements of Prior Learning Experience and demographic information were collected and coded before the beginning of the class; after the 32-week (2 semesters) FCM sample course, the participants were asked to complete the rest of the questionnaires. Totally 158 students answered the questionnaires anonymously, of which 146 valid questionnaires were utilizable since 12 responses were incomplete and had to be discarded, giving a valid response rate of 92.41% of the initial sample. Although the number of participants was limited, the results (shown in Table 1) of a chi-square test indicated that in terms of gender (p = .19), major (p = .24), grade (p = .11), and spending time online (p = .75), there were no statistically significant differences in the demographics, suggesting that this sample is relatively representative.

The current FCM teaching experiment was carried out from September 2013 to June 2014. Figure 2 shows that this course consisted of three parts (before class, in class and after class) with six 45-minute classes each week, which is in accordance with class periods in traditional settings. Three physical classes (40-50 students per class, 3 classes in total) were carried out at a set time and place each week, where the students were divided into groups (5-6 students per group in random) according to the seats settings in the physical classroom for the convenience of discussion, and online preview and revision took up the remaining three classes depending on the learners’ personalized arrangement with the online learning platform (shown in Figure 3). The platform consisted of two basic learning functions: video platform and discussion platform. To increase the storage capacity, the lecture videos captured by instructors are uploaded to a popular video website You Ku, to which the URL was
linked to the platform. Besides, the discussion platform including BBS and online chat room are both designed for learners’ convenient mind-exchanging.

Table 1. Demographic profile and results of chi-square test (n = 146)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Classification</th>
<th>Total (%)</th>
<th>( \chi^2 ) (Sig.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>81 (0.55)</td>
<td>1.75 (0.19)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>65 (0.45)</td>
<td></td>
</tr>
<tr>
<td>Major</td>
<td>Engineering</td>
<td>41 (0.28)</td>
<td>4.19 (0.24)</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>44 (0.30)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arts</td>
<td>29 (0.20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>32 (0.22)</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>Freshman</td>
<td>41 (0.28)</td>
<td>6.06 (0.11)</td>
</tr>
<tr>
<td></td>
<td>Sophomore</td>
<td>32 (0.22)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Junior</td>
<td>27 (0.18)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Senior</td>
<td>46 (0.32)</td>
<td></td>
</tr>
<tr>
<td>Time Online</td>
<td>&lt;3h</td>
<td>33 (0.23)</td>
<td>1.23 (0.75)</td>
</tr>
<tr>
<td>(1 week)</td>
<td>3–6h</td>
<td>42 (0.29)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6–9h</td>
<td>36 (0.25)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;9h</td>
<td>35 (0.24)</td>
<td></td>
</tr>
</tbody>
</table>

Data analysis

Several statistical indices need to be considered when conducting data analysis. Cronbach alpha coefficients and composite reliabilities for all constructs were first examined using SPSS 19 to evaluate the internal reliability of the nine constructs. Then the correlation analyses were presented for examining the relationships between each construct. Furthermore, Structural Equation Modeling (SEM) conducted using AMOS 22.0 with the method of maximum likelihood estimation was applied to achieve the purposes of this study. As Stevens (2012) suggested, a suitable sample case when conducting CFA could be five to ten times of the total number of items in the surveying scales; the current survey with 146 cases satisfies this criterion. The structural model was tested by examining the relationships among the latent variables and detecting the fitness of the proposed models. Thus, SEM was considered to be an appropriate technique for the current study. Specifically, several model fit statistics were tested, including the ratio of chi-square to the degree of freedom, RMSEA (the root mean square error of approximation), CFI (comparative fit index), IFI (incremental fit index), and GFI (adjusted goodness of fit index). Besides, path coefficients were analyzed to test each hypothesis.
Instruments

To guarantee the validity of the measurements, a large number of prior relevant studies were consulted to ensure that a comprehensive list of measurements was included, and appropriate instrument items could be employed from past studies. The instruments of Prior Learning Experience adapted from Bourgonjon, Valcke, Soetaert, and Schellens’s (2010) research (loading: 0.52-0.88) investigated whether learners were experts in blended learning before the beginning of the FCM trial class. Additionally, three items modified from Paechter and Maier’s study (2010) were tapped to measure the students’ perceived extent of Personalized Learning Climate in FCM. Perceived Value was measured with items adapted from Fornell et al.’s survey (1996), and in the FCM context, learners’ consumed time rather than charged expense was considered as learners’ “cost.” To evaluate students’ satisfaction with FCM, items comprehensively adapted from Chen, Lin, and Kinshuk’s study (2008) (loading: 0.64, 0.88 respectively) and Wu, Tennyson, and Hsia’s study (2010) (composite reliability coefficients: 0.957; the AVE: 0.849).

The perceived quality instrument including five factors was mainly adapted from prior relevant studies, which have indicated the satisfactory reliability of each factor. Items for perceived usefulness of the platform (UP) and ease of using the platform (EUP) were adapted from Tsai et al.’s work (2012), and the reliability (Cronbach’s alpha) for UP and EUP were 0.84 and 0.93, respectively, indicating that these factors had sufficient reliability. Besides, Usefulness of the online course (UOC) and Ease of using the online course (EUOC) were measured with items from Sun et al.’s study (2008). The reliability of the measured items in Sun’s study was respectively 0.87 and 0.83 for the UOC and EUOC. Additionally, the Interaction in physical classes (IPC) items in this survey were adapted from Paechter and Maier’s study (2010) to investigate the students’ evaluative standards in the situations of judging peers and instructor-pupil interactive effectiveness in a face-to-face classroom.

Four professors (two majoring in education technology and two in foreign language) validated the questionnaire items by individual review in paper-and-pencil format and group discussion when translating the content of the items into Chinese, and some items were consolidated and modified in response to the FCM context. Besides, a pilot survey was conducted with 10 students, then the format and content of the questionnaire were refined accordingly. Consequently, a total of 26 items presented with a 5-point Likert scale (from 1, “strongly disagree” to 5, “strongly agree”) were administered to investigate the university students’ satisfaction with the flipped EFL course.
Results

Correlation analysis

The means and standard deviations of each variable as well as their correlations are shown in Table 2. The “SS” was significantly and positively correlated with all of the variables, which provided initial suggestions of the relationship between “SS” and other variables. Specifically, the correlation between “SS” and “PLE,” “PV” was relatively higher ($r = 0.51, 0.58$ respectively), while the correlation between “SS” and “PLC” was relatively lower ($r = 0.37$), suggesting that “PLE” and “PV” may play a more vital role in driving students’ perceptions in FCM. However, there was no significant correlation between “PLC” and “PV,” which generally indicated that in the flipped context the personalized settings may not directly relate to students’ perceived value.

Table 2. Means, standard deviations and correlations of the research variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.PLC</td>
<td>3.85</td>
<td>0.65</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.PLE</td>
<td>3.61</td>
<td>0.68</td>
<td>0.72</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.PV</td>
<td>3.77</td>
<td>0.78</td>
<td>0.70</td>
<td>0.72</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.SS</td>
<td>3.79</td>
<td>0.74</td>
<td>0.71</td>
<td>0.70</td>
<td>0.72</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.EUP</td>
<td>3.72</td>
<td>0.64</td>
<td>0.70</td>
<td>0.71</td>
<td>0.70</td>
<td>0.72</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.UOC</td>
<td>3.61</td>
<td>0.71</td>
<td>0.70</td>
<td>0.71</td>
<td>0.70</td>
<td>0.72</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.IPCC</td>
<td>3.56</td>
<td>0.72</td>
<td>0.70</td>
<td>0.71</td>
<td>0.70</td>
<td>0.72</td>
<td>0.71</td>
<td>0.72</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. $n = 146$. * $p < .05$, ** $p < .01$. Personalized Learning Climate (PLC), Prior Learning Experience (Baeppler et al.), Perceived Value (PV), Student Satisfaction (Baeppler et al.), Ease of using the platform (EUP), Usefulness of the platform (UP), Ease of using the online course (EUOC), Usefulness of the online course (UOC), Interaction in physical classes (IPC).

CFA analysis

Confirmatory factor analysis was performed to examine the measurement model of the FCM students’ satisfaction index. According to Table 3, all of the factor loadings are greater than 0.50, and the $t$-values revealed significance at the 0.001 level, indicating a strong relationship with their associated constructs. Additionally, the Cronbach alpha coefficients for each scale ranged from 0.64 to 0.80 (overall alpha = 0.82), and each of the composite reliability values was above 0.50, which revealed the acceptable internal consistency of this questionnaire. The fitness indices of the measured items for the proposed model (the ratio of chi-square to degrees of freedom = 1.18, RMSEA = 0.035, GFI = 0.91, IFI = 0.96, CFI = 0.96) indicated an acceptable model fit. Meanwhile, the squared multiple correlation of Students Satisfaction presented in parentheses was 0.73, indicating a good examination of students’ perceived satisfaction. Moreover, path analysis was employed to assess the structural model that specified the relations between the latent constructs. In sum, the above reliability and validity of the data coefficient showed that the measurement was acceptable.

Table 3. CFA analysis for students satisfaction questionnaire in FCM ($n = 146$)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item</th>
<th>Loading</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of using the online course (EUOC)</td>
<td>Source: (Sun, Tsai, Finger, Chen, &amp; Yeh, 2008)</td>
<td>--------</td>
<td>0.63</td>
</tr>
<tr>
<td>1 Learning to operate the online course system was easy for me.</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Generally, I find that the online course system is easy to use.</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 It is easy to get the online course system to do what I want it to do.</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 It was easy for me to become skillful at using the online course.</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usefulness of the Online course (UOC)</td>
<td>Source: (Sun, Tsai, Finger, Chen, &amp; Yeh, 2008)</td>
<td>--------</td>
<td>0.73</td>
</tr>
<tr>
<td>1 Taking the online course improved my effectiveness in the FCM program.</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Using the online course in the FCM program enhanced my productivity.</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 I feel the usefulness of conducting flipped classes via the online course.</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Quality</td>
<td>Ease of using the platform (EUP)</td>
<td>Source: (Tsai, Hwang, Tsai, Hung, &amp; Huang, 2012)</td>
<td>--------</td>
</tr>
<tr>
<td>1 It took only a short time to learn how to operate the FCM interactive system.</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2 The design of the FCM platform interface fits users’ operating habits.
3 It is easy and clear for me to search needed results in the FCM platform.

**Usefulness of the platform (UP)**

Source: (Tsai, Hwang, Tsai, Hung, & Huang, 2012)

<table>
<thead>
<tr>
<th>Source</th>
<th>UP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.63</td>
</tr>
</tbody>
</table>

2 The use of the interactive platform enhanced my ability to search for information when problem-solving.
3 The use of the interactive platform made me have more interest in the course.

**Interaction in physical classes (IPC)**

Source: (Paechter & Maier, 2010)

<table>
<thead>
<tr>
<th>Source</th>
<th>IPC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.73</td>
</tr>
</tbody>
</table>

1 I can effectively exchange information in the FCM physical class.
2 I can get support from cooperative learning and group work with other participants in the FCM physical class.
3 I can easily get counseling and support by the tutor in the FCM physical class.

**Personalized Learning Climate (PLC)**

Source: (Paechter & Maier, 2010)

<table>
<thead>
<tr>
<th>Source</th>
<th>PLC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.76</td>
</tr>
</tbody>
</table>

1 I can decide on my own the time and pace for FCM learning.
2 It is flexible with regard to FCM learning strategies.
3 The flipped classroom model provides personalized learning support.

**Prior Learning Experience (PLE)**

Source: (Bourgonjon, Valcke, Soetaert, & Schellens, 2010)

<table>
<thead>
<tr>
<th>Source</th>
<th>PLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.64</td>
</tr>
</tbody>
</table>

1 Compared to people of my age, I participate in a lot of blended learning activities (E-learning + face to face strategy).
2 Compared to people of my age, I am in favor of participating in blended learning activities.

**Perceived Value (PV)**

Source: (Fornell et al., 1996)

<table>
<thead>
<tr>
<th>Source</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.80</td>
</tr>
</tbody>
</table>

1 Compared with the traditional learning approach, I spend less time learning a certain knowledge point in FCM.
2 Compared with the traditional learning approach, I can learn more in a certain time in the flipped context.

**Students’ Satisfaction (SS)**

Source: (Chen, Lin, & Kinshuk, 2008; Wu, Tennyson, & Hsia, 2010)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.77</td>
</tr>
</tbody>
</table>

1 Overall, I feel satisfied with flipped pedagogy.
2 I am satisfied that FCM meets my needs in terms of learning.
3 I would like to continually use the FCM in my learning.

**SEM analysis**

The SEM standardized path analysis results for testing each hypothesis are shown in Figure 4. “Personalized Learning Climate” and “Previous Learning Experience” are positively associated with “Student Satisfaction” (path coefficient 0.34, 0.45, respectively, p < .01, p < .01, respectively), which confirms hypotheses H1 and H2. “Personalized Learning Climate” is positively associated with “Perceived Quality” (path coefficient 0.38, p < .01) which confirms hypothesis H3c. Besides, “Prior Learning Experience” has significant positive relationship with “Perceived Quality” and “Perceived Value” (path coefficient 0.52, 0.40 respectively, p < 0.01), which supports hypotheses H3b and H4a. “Perceived Value” is positively associated with “Student Satisfaction” (path coefficient 0.51, p < .01), which supports hypothesis H8. Furthermore, “Perceived Quality” significantly fosters “Perceived Value” (path coefficient 0.49, p < .01) which supports H5. Contrary to our predictions, “Perceived Quality” fails to directly predict “Student Satisfaction”, and “Personalized Learning Climate” is not significantly related to “Perceived Value”; therefore, hypotheses H3a and H4b are unsupported.

**Discussion**

The results of this survey indicated that prior learning experience and personalized learning climate are both important antecedents to predict students’ perceptions in FCM settings. Personalized settings allowing learners to work at their own pace are helpful for increasing student satisfaction, which is also consistent with Chow and Shi’s (2014) argument.
Moreover, compared with the design of a personalized learning climate, learners’ prior learning experience is a far more significant predictor of learners’ favorable perceptions in the flipped context, and the efforts to strengthen learners’ prior experience may be more “economical” and operable. Accumulated experience of blended learning fosters students’ comprehension of learning procedures, benefits, and learning difficulties, it leaves learners with a clear judgment of their preference for creating student-centered instruction for themselves.

![Diagram of structural relationship in FCM](image)

**Notes.** *p* < .01. Numbers in parentheses are squared multiple correlations.

Figure 4. The final model of structural relationships in FCM

Surprisingly, contrary to our hypothesis, personalized learning climate could not directly predict students’ perceived value, and perceived quality completely rather than partially mediated personalized learning climate and perceived value. The possible explanation for H4c being unsupported may be the fact that the personalized climate in FCM provides learners with a free and personalized space. However, like a double-edged sword, in such a “free” climate, students’ attention is easily distracted by computer games or online free chat, if the course quality is inadequate to attract them, and this is particularly so for those students whose autonomous learning capability tends to be relatively weak. Therefore, it is suggested that the superior course quality still plays a vital role in mediating the two factors.

Unexpectedly, the influence of perceived quality on student satisfaction is weak, indicating that perceived value is a complete mediator between perceived quality and students’ satisfaction. The fact that H3a was unsupported highlights the fact that perceived quality was inclined to predict students’ satisfaction through the mediator of “perceived value,” which draws special emphasis on learning efficiency in FCM settings. Although learners could watch the videos repeatedly and speculate on any determined points, they might still feel stressed and confused. Because they failed to receive peers/teachers’ definite responses instantly, until they attended physical classes a few days later. Thus, learners’ perceived quality is necessarily related to their satisfaction through the complete mediator perceived value.

**Conclusion**

This study, based on ELT, adapted developed instruments and utilized SEM to empirically examine student satisfaction in a flipped classroom context. We believe that our theoretical development provides a step toward a better understanding of FCM effectiveness.

The findings of this research thus raise vital implications for FCM teaching and learning practice. First and foremost, the “pre-instruction” is of greater significance for learners’ in-depth understanding of FCM design and procedure. Thus a series of warm-up activities as “rehearsals” are suggested carrying out for students’ comprehension of their mission before/in/after the class, which is followed by learners’ favorable perception of their self-confidence, interest points and learning controllability. Specifically, instructors should offer opportunities for learners to observe the flipped procedures at least through viewing some demonstration video on flipped learning to help learners fully understand the FCM procedures and practices before the start of the formal course. Some training classes concerning the usage of the online platform ought to be performed to illustrate the details of using the platform. Besides, according to the scale of PLE, learners’ previous learning experience has potential relationship with their preference of the blend learning model. Therefore, instructors have to take learners’ individual difference into consideration when conducting FCM. For instance, it can be adaptive for meeting the needs of students with difference levels of maturation or preferences for blended learning methods. Finally, considering perceived value is conducted as a complete mediator in the satisfaction
model, some intelligent techniques available for enhancing learning efficiency, such as instant online chat apps, intelligent agents, are strongly suggested being utilized when designing the online learning platform for flipped learning.

Although a rigorous validation procedure was conducted to develop a general instrument for measuring FCM satisfaction, this research still suffers from the limited sample and course type in the current survey, since flipped pedagogy is still in its infancy in China. To deepen our understanding of students’ perceptions, future research is therefore warranted to analyzing factors considering the distinction of course types and learners’ personality with the growing popularity of flipped pedagogy.

Acknowledgments

Thanks are due for funding by the National Natural Science Foundation of China (NO.: 71371177 & NO.: 61300060), the Ministry of Education Project of Humanities and Social Sciences (MOEPHSS: 13YJA880020), and Anhui provincial research projects (foundation NO.: 2015zdjy115, 2015zdjy206 & NO.: SK2015A632).

References


Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. Eugene, OR: International Society for Technology in Education.


Implementing the Flipped Classroom in Teacher Education: Evidence from Turkey

Gökçe Kurt
Department of English Language Teaching, Marmara University, Istanbul, Turkey // gokce.kurt@marmara.edu.tr

ABSTRACT
The flipped classroom, a form of blended learning, is an emerging instructional strategy reversing a traditional lecture-based teaching model to improve the quality and efficiency of the teaching and learning process. The present article reports a study that focused on the implementation of the flipped approach in a higher education institution in Turkey. For this pretest-posttest quasi-experimental study, a classroom management course in a pre-service English teacher education program was flipped and its effectiveness was measured against a traditionally taught class. Quantitative and qualitative data came from 62 pre-service teachers (PTs) in two intact classes randomly assigned as the experimental and the control groups. Findings revealed a higher level of self-efficacy beliefs and better learning outcomes for the experimental group PTs in the flipped classroom compared to the control group PTs in the traditional classroom. PTs’ perceptions of the flipped classroom were also positive.

Keywords
Flipped classroom, Teacher education, Classroom management, Pre-service teachers of English

Introduction
Enhancing students’ learning experiences and meeting their needs and expectations have been among the primary concerns of higher education institutions over three decades (Demirer & Sahin, 2013; Garrison & Hanuka, 2004). Therefore, extensive research has been conducted in order to investigate the ways of improving the teaching and learning environment (Biggs & Tang, 2011). Blended learning (BL) has emerged with its “potential to transform higher education institutions” to supplement the traditional classroom setting and increase the quality of student learning (Garrison & Hanuka, 2004).

Despite its growing popularity, there is not an agreed upon single definition of BL in the literature. In Singh and Reed’s (2001) definition, BL refers to a learning program using more than one delivery model to increase students’ academic achievement and reduce costs. In Bañados’ words (2006), BL is “a combination of technology and classroom instruction in a flexible approach to learning that recognises the benefits of delivering some training and assessment online but also uses other modes to make up a complete training programme which can improve learning outcomes and/or save costs” (p. 534). According to Garrison and Vaughan (2008), BL is about the thoughtful integration of face-to-face and online learning; the redesigning of a course to increase student engagement; and the restructuring and replacement of traditional class hours. They add that BL blends “the best of traditional and Web-based learning experiences,” and this blend or integration is “multiplicative, not additive” (p. 7). Common to these definitions is the recognition of BL as the combination of face-to-face and online-delivery methods to enhance students’ learning and reduce costs.

The use of BL has been reported to be on the rise in higher education (Garrison & Vaughan, 2008; Graham, Woodfield, & Harrison, 2013). Its popularity is attributed to research suggesting that BL results in improved learning outcomes (Bonk & Graham, 2006; Vaughan, 2010), improved pedagogy (Graham, 2006), increased learner motivation and satisfaction (Collopy & Arnold, 2009; Fulkerth, 2010), increased interactions (Bonk & Graham, 2006; Delialioglu & Yildirim, 2007), and reduced costs (Bonk & Graham, 2006; Graham, 2006).

In recent years, the flipped or inverted classroom, falling under the broad definition of blended learning, has gained prominence in higher education institutions as an alternative pedagogical model reversing what traditionally occurs in and out of the class.

The flipped classroom
The flipped classroom is an emerging pedagogical model in which traditional lecture is moved outside the classroom via technology and assigned as homework while in-class time is spent on collaborative inquiry-based learning (Bergman & Sams, 2012; Lage, Platt, & Treglia, 2000; Stone, 2012; Tucker, 2012). The flipped classroom is defined by Bishop and Verleger (2013) as follows:
The flipped classroom is a new pedagogical method, which employs asynchronous video lectures and practice problems as homework, and active, group-based problem solving activities in the classroom. It represents a unique combination of learning theories once thought to be incompatible- active, problem-based learning activities founded upon a constructivist ideology and instructional lectures derived from direct instruction methods founded upon behaviourist principles. (p. 1)

The flipped classroom model has developed through the years with contributions from several researchers. In 1996, Mazur argued against using lectures and having passive students in the classroom. He used a technique called “Peer Instruction” to change the traditional instructional model. He asked his students to prepare for class by reading the assigned materials so that they could be actively engaged in the learning process with their peers (Mazur, 2009). Similarly, Baker (2000) shared lecture notes with his students online and used class time for group work and problem solving activities. In the same year, Lage et al. (2000) also inverted their classroom and provided students with a choice of videotaped lectures, reading materials and slides to study outside the class. In-class time was spent on group work activities based on practice. In 2006, Salman Kahn launched his non-profit website offering free videos on almost every topic to support instruction. In 2007, with the study of Bergmann and Sams (2012), the flipped model became more popular. They shared recorded teacher presentations online to support students who were frequently absent. Although this was their initial purpose, in time, their study showed that the flipped classroom resulted in greater teacher-student rapport and increased student-student interaction. More in-class time was allocated to conducting engaging activities.

As a student-centred instructional model, the flipped approach is grounded in the constructivist theory of learning (Strayer, 2012). In constructivism, “knowledge is actively constructed by the learner, not passively received from the outside. Learning is something done by the learner, not something that is imposed on the learner” (Sjoberg, 2007, p. 3). The flipped approach aims to create a student-centred learning environment in which students take care of their own learning and become more active and interactive in class. Among the foundations of constructivist theory of learning are collaboration, interaction and engagement of learners (Jonassen, Davidson, Collins, Campbell, & Haag, 1995). In the flipped classroom, students “work through problems, advance concepts, and engage in collaborative learning” (Tucker, 2012, p. 82). Such collaborative activities are effective in supporting students’ higher level of understanding (Sorden, 2011; Yang & Wu, 2012).

Differentiated instruction is also evident in the flipped classroom. Differentiating, or personalising, instruction involves the identification of the needs and preferences of learners and the organisation of instruction that is meaningful and relevant to their learning (Algozzine & Anderson, 2007; Keefe, 2007). In the flipped approach, students have the opportunity to learn according to their needs and preferences. They can, for example, rewind or fast-forward to review the materials shared prior to class and decide on their own pacing. In the classroom, teachers have more time to guide students and give them differentiated feedback (Berrett, 2012).

**Research on the flipped classroom**

Despite its rapid rise in popularity in higher education, there are relatively few studies on the efficacy of the flipped model. Stone (2012) flipped a class of a genetic diseases course with 30 students and a class of a general biology course with 400 students. In both courses, students were provided with short recordings of lectures, readings, animations and simulations prior to the lesson. In class, activities such as jigsaw exercises and case studies were conducted. Compared to students in previous non-flipped semesters, the flipped class students scored higher on exams, attended the classes more and had mostly positive attitudes towards learning in the flipped classroom. The study conducted by Marcey and Brint (2012) in two introductory biology classes aimed to compare a traditional lecture model with the flipped classroom approach. At the end of the study, students in the flipped classroom had higher scores on tests and quizzes than the students in the lecture class. Davies, Dean and Ball (2013) compared three different teaching approaches, i.e., traditional lecture-based instruction, simulation-based instruction and the flipped classroom in a course on spreadsheets with 301 students. The results indicated the effectiveness of the flipped classroom. Due to its allowance for greater differentiation of instruction, this approach increased students’ motivation and facilitated their learning. Similarly, in their study, McLaughlin et al. (2014) flipped a pharmaceutical course with 162 students by creating online videos of their lectures and using class time for active learning exercises. The findings revealed an increase in students’ learning and their perceived value of the model.

Compared to research on the implementation and efficacy of the flipped classroom approach in other disciplines, there is very little research on the effectiveness of the model in teacher education specifically; furthermore, the findings are inconclusive. In one such study, Vaughan (2014) investigated the effect of a flipped classroom
approach on PTs’ engagement level in a teaching profession course. Data came from discussion board notes, journals and observations. During the study, PTs in the flipped classroom watched the taped lectures and answered discussion questions outside the classroom and participated in activities such as small group debates inside the classroom. The findings revealed an increased level of reflection and inquiry for the PTs in the flipped group. Fraga and Harmon (2014) also aimed to investigate PTs’ perceptions of the flipped classroom and the effect of the approach on their achievement by comparing the flipped and the traditional lecture classrooms in a word study course. Fifty-one PTs enrolled in the Department of Education and Human Development participated in the study. The flipped group PTs were provided with lectures in the form of narrated PowerPoint presentations. The findings did not reveal any significant difference between the groups in terms of PTs’ academic achievement. The authors noted that different factors such as topic choices, preferences and learning styles of PTs influenced the effectiveness of the way teaching and learning was conducted.

The present study aimed to contribute to the literature on the implementation and effectiveness of the flipped approach in pre-service English teacher education by comparing a flipped classroom to a traditional lecture-based classroom. More specifically, a course titled “Classroom Management” was flipped and this study was conducted to investigate whether there would be significant differences between the PTs of English enrolled in the flipped classroom management course and those enrolled in the traditional face-to-face classroom management course in terms of their self-efficacy beliefs and learning outcomes. PTs’ perceptions of the flipped classroom approach were also analysed.

**Methodology**

**Design, participants and setting**

This study adopted a pretest-posttest quasi-experimental mixed methods design. The experiment was conducted over a 14-week semester with 62 PTs of English enrolled in the four-year English Language Teaching program of a state university in Turkey. They were all second-year students and took a 4-credit Classroom Management course as a curriculum requirement. At the time of the study, the course was offered in four sections. Among these four intact classes, two of them were randomly selected for the study and assigned to control ($N = 30$) and experimental ($N = 32$) conditions, receiving traditional and flipped instruction respectively. All sections of the course were taught by the same instructor, the researcher of this study.

Demographic characteristics of the participants were as follows: the majority of the PTs in both groups were female (80% in the control group, 81% in the experimental group). Their ages ranged from 19 to 21. Every PT except one in the control group owned a personal computer and had Internet access. Eighty percent of the PTs in the control group spent up to five hours a day online while the rest reported to be spending five to 10 hours. Similarly, 81% of the experimental group PTs spent about five hours a day on the Web while 19% were spending up to 10 hours. All PTs reported that they used the Internet mainly for social media, communication, entertainment, and education. Their educational uses of the Internet included using an online dictionary, searching the Web for a school project, using Wikipedia, using Edmodo, and practicing language. None of the PTs had heard about blended learning or the flipped approach prior to the study.

**The classroom management course and its redesign**

Classroom management is considered to be one of the most important skills that teachers and PTs gain for effective learning and teaching (Stoughton, 2007). It is also commonly reported to be a cause of concern for PTs during their practicum and for their future teaching career (Bromfield, 2006). Most teacher preparation programs offer courses on classroom management (Baker, 2005). These courses generally focus on the development of theoretical knowledge through lecturing and opportunities for putting theory into practice are rare. Research focusing on the views of PTs for the adequacy of classroom management preparation during teacher education has revealed that they do not feel adequately prepared, they feel a need for additional training in classroom management, and find the courses too theoretical, not connected to the real world (Atay & Kurt, 2010; Atici, 2007; Maskan, 2008).

The Classroom Management course that is highlighted in this study is conducted in English and aims to prepare PTs to be effective managers of their future classes to maximise learning in the classroom environment. Key topics of study for this course include the development of a favourable and cooperative climate in the classroom, the causes of discipline problems in the classroom and the strategies for dealing with them, and the development
of effective relationships among students, teachers and parents. The course content is traditionally covered by lectures. Students are required to read the assigned chapters before coming to class and follow the lectures during the lesson. Classroom interaction is mainly based on teacher-led question-and-answer sequences. Due to limited class time and crowded classes containing around 30-35 PTs, only few PTs get the opportunity to talk and not much time is left for practical activities.

Prior to the present study, informal discussions with PTs who had already taken the course revealed their concern for classroom management. They felt unprepared to deal with classroom management problems in a real classroom setting, even after taking the course. They considered the lack of opportunity to practice to be the most significant factor contributing to their low self-confidence regarding classroom management. They also mentioned that as the course was mostly theoretical, following lectures could sometimes be difficult and boring.

For the above mentioned reasons, one section of the classroom management course was redesigned, or flipped, in order to (1) create time for activities that would help PTs put their theoretical knowledge into practice, (2) make lessons more interactive and engaging, and (3) improve PTs’ learning. The other section assigned as the control group continued with the traditional format.

**Description of the flipped instruction**

Utilizing the Constructivist theory of learning and differentiated instruction in its design, the flipped Classroom Management course was conducted as follows: At the beginning of the term, a group was created on Edmodo (http://www.edmodo.com), a social learning platform for teachers, students and schools. Edmodo allows teachers to create groups for their classes and post classroom materials, share links and videos, and create quizzes and assignment tasks. The PTs were familiar with Edmodo; thus, there was no need for training on its use.

For the content delivery in the flipped group, lectures of the course were recorded as video podcasts (vodcasts) using Present.me (https://present.me/content/), an online service that allows users to combine slides with audio and video narration and create links for viewers. Each week, an instructor-generated vodcast, typically 40-45 minutes in length, was created and its link was shared on Edmodo. For each vodcast, an online quiz was prepared on Edmodo in order to check PTs’ comprehension of the lecture. PTs were supposed to come to class having viewed a lecture vodcast linked on Edmodo, read the assigned chapter(s) in the textbook, and completed the quiz. PTs were reminded to take notes of any questions that arose during their self-study.

Face-to-face class time of the flipped group began with a brief review of the lecture video content. In the remaining time, approximately 40-45 minutes, practice-based interactive tasks were carried out, mostly in pairs or groups. Tasks ranged from problem solving to analysis of case scenarios, from role-playing to watching and commenting on real classroom videos. The aim was to give PTs the opportunity to grapple with real-world problems and apply theory to practice. For example, in a week focusing on different approaches to classroom management, i.e., the Jones Approaches, the Dreikurs Approach or the Behaviourist Approach, PTs learned about the relevant principles, facts and terms from the lecture vodcasts out of the classroom. In-class time was spent on analysing case scenarios that may be potentially faced in a real classroom and working on possible responses based on the approaches covered in the video. More specifically, PTs worked collaboratively in groups of three or four to act out the scenarios they were assigned to and present their responses. After each presentation, whole class discussions were carried out to discuss the applicability of PTs’ responses in a real classroom setting. In another week, the focus was on communicating with students and their parents. PTs came to class having learned strategies for successful teacher-student and teacher-parent communications from the assigned vodcast and readings. During the face-to-face class time, PTs worked with their peers and role-played a teacher-parent interaction during a teacher-parent meeting. They were given role cards and expected to follow the conventions for successful communication in teacher-parent meetings.

**Description of the traditional instruction**

The instruction in the control group was built around in-class lectures. A PowerPoint presentation was used for each lecture focusing on the theory. The content of the presentations used in the flipped and the traditional groups was the same. PTs in the traditional group were supposed to come to class having read the assigned chapters in the textbook. As homework, they were also asked to evaluate teacher responses to classroom management scenarios presented in each chapter and suggest alternative strategies. Due to the limited time and
large class size, not much time could be allocated to the review of the scenarios and only few PTs got the opportunity to share their opinions and/or suggestions.

**Data collection and analysis procedure**

For the present study, quantitative data came from the Teachers’ Sense of Efficacy Scale (TSES) developed by Tschannen-Moran and Hoy (2001) and PTs’ final exam scores. Qualitative data were obtained from a focus group interview.

On the first day of class in both groups, the syllabus was reviewed and course requirements were discussed. The flipped group was informed about the flipped approach and instructed on how to access the vodcasts linked to the group page on Edmodo. The first in-class meeting in each group ended with the administration of the TSES. After completion of the course content and before the final exam, the same scale was applied to both groups.

The TSES, with 24 items grouped in three subscales, aimed to measure PTs’ perceived self-efficacy in student engagement, instructional practices and classroom management. Example items for three subscales were as follows: “How much can you assist families in helping their children do well in school?” (student engagement), “To what extent can you use a variety of assessment strategies?” (instructional strategies), and “How much can you do to get children to follow classroom rules?” (classroom management). PTs indicated how much they could do in the given statements on a 9-point Likert scale ranging from 9 (a great deal) to 1 (nothing). The Cronbach alpha reliability coefficient of the overall scale was .94, and the reliability scores for three subscales were .87 for student engagement, .91 for instructional strategies, and .90 for classroom management (Tschannen-Moran & Hoy, 2001).

For data analysis, first, descriptive statistics were conducted using SPSS 21.0 to investigate the overall self-efficacy levels of PTs. Then, an independent samples t-test was used to compare the efficacy levels of the PTs who received the traditional lecture-based classroom management course to those who received the flipped instruction.

Upon completion of all course content in both groups, a final exam consisting of 20 fill-in-the-blank questions was administered to the PTs in both the control and experimental groups. An independent samples t-test was used to determine whether the groups’ learning outcomes differed significantly from one another. The exam included questions measuring both theoretical knowledge and decision making based on given cases.

Qualitative data came from a focus group interview conducted with nine randomly selected PTs from the flipped group. The aim of the interview was to understand PTs’ perceptions and experiences of the flipped instruction after receiving it. The interview, which lasted around 40 minutes, was carried out in the PTs’ native language (Turkish) and audio recorded for transcription purposes. The transcribed data were translated into English and analysed by repeated reading to identify and report common patterns/themes and divergences within the data, following the procedure of Miles and Huberman (1994).

**Findings**

**Teachers’ sense of efficacy scale**

In the preliminary analyses, the assumption of normality was assessed with the Kolmogorov Smirnov (KS) test using SPSS software. The KS test results for the pre- and post-test scores were not significant, indicating that the data came from a normally distributed population, verifying the assumption ($p = .20$, for both).

In order to investigate self-efficacy beliefs of PTs, an independent samples t-test was applied to both groups’ gain scores in the TSES. As Table 1 shows, there was a significant difference between the experimental and control groups in their gain scores in the overall scale and the subscales of student engagement and classroom management. For the subscale of instructional strategies, PTs in the experimental group had higher gains compared to the PTs in the control group, though not at a significant level. As the eta squared values indicate, the magnitude of the differences in the means obtained from the overall scale, and the subscales of student engagement and classroom management was moderate (Cohen, 1988).
Table 1. Differences in self-efficacy beliefs

<table>
<thead>
<tr>
<th>Scale</th>
<th>Group</th>
<th>Test</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>ŋ²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>Pre</td>
<td>42.41</td>
<td>9.51</td>
<td>60</td>
<td>2.845</td>
<td>.006*</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>(N = 32)</td>
<td>Post</td>
<td>54.59</td>
<td>7.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Pre</td>
<td>45.70</td>
<td>7.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N = 30)</td>
<td>Post</td>
<td>51.70</td>
<td>6.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructional</td>
<td>Pre</td>
<td>41.69</td>
<td>10.80</td>
<td>60</td>
<td>1.343</td>
<td>.184</td>
<td></td>
</tr>
<tr>
<td></td>
<td>strategies</td>
<td>Post</td>
<td>52.97</td>
<td>6.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>Pre</td>
<td>43.27</td>
<td>8.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N = 32)</td>
<td>Post</td>
<td>50.93</td>
<td>7.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Pre</td>
<td>47.10</td>
<td>8.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N = 30)</td>
<td>Post</td>
<td>53.23</td>
<td>6.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>Pre</td>
<td>41.13</td>
<td>10.65</td>
<td>60</td>
<td>2.745</td>
<td>.008*</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>management</td>
<td>Post</td>
<td>55.16</td>
<td>7.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental</td>
<td>Pre</td>
<td>47.10</td>
<td>8.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N = 30)</td>
<td>Post</td>
<td>53.23</td>
<td>6.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Pre</td>
<td>48.47</td>
<td>8.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N = 30)</td>
<td>Post</td>
<td>54.93</td>
<td>7.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Pre</td>
<td>125.22</td>
<td>27.30</td>
<td>60</td>
<td>2.522</td>
<td>.014*</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>Post</td>
<td>162.72</td>
<td>21.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05. ŋ² = Eta squared.

Final exam scores

The study also focused on understanding whether there was a significant difference between the experimental and control group PTs in terms of their exam scores at the end of the study. The preliminary analysis using the KS test revealed that the scores were normally distributed (p = .089); therefore, an independent samples t-test was applied to the data. The findings of the t-test displayed that (see Table 2) PTs in the flipped classroom outperformed PTs in the traditional classroom and the difference was statistically significant. The eta squared statistic indicated a large effect size (ŋ² ≥ .14, Cohen, 1988).

Table 2. Differences in final exam scores

<table>
<thead>
<tr>
<th>Group</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>ŋ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>73.38</td>
<td>14.64</td>
<td>60</td>
<td>3.716</td>
<td>.000*</td>
<td>.19</td>
</tr>
<tr>
<td>(N = 32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>58.80</td>
<td>16.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05. ŋ² = Eta squared.

Focus group interview

A focus group interview was conducted in order to gain an in-depth understanding of the experimental group PTs’ perceptions of and experiences with the flipped instruction. Four major categories emerged from the analysis of the data: perceptions of the efficacy of the flipped classroom model, changes in the classroom environment, benefits of individualised learning, and applicability of the model to language teaching.

The first theme emerging from the data was about PTs’ positive perceptions of their experiences in the flipped learning environment. All PTs reported to be highly satisfied with the method of instruction they received. Two major factors that led to satisfaction were their perception of better learning and their enjoyment of the flipped class model. They stated that watching the videos prior to in-class meetings helped them learn the material better and was more enjoyable compared to doing assigned reading. One PT said:

*I feel more motivated to watch the lecture from a video. Reading a chapter or an article is usually boring. I learn better from the videos. During the term, I enjoyed attending the class knowing the material and feeling ready for discussion.*

Another PT stated that:

*Most courses in our program are theoretical. When the focus is on theory, I memorize the content and forget it easily. In this course, I learned the theory at home and we had a lot of practice in class. This way, I could connect theory and practice. They were complementary. I feel I can confidently put my theoretical knowledge into practice in a real classroom environment.*
In relation to the efficacy of the model, the following comment came from another PT:

*I really enjoyed the course, especially our in-class meetings. I felt that I was learning. Activities such as case analysis and role-playing contributed to our learning and prepared us well for a real classroom. I will remember what I have learned in this course in the future.*

PTs’ comments also focused on the classroom environment. They mentioned that in the flipped approach, the classroom environment was student-centred, more positive and less stressful. One PT stated the following about the student-centred learning environment:

*In a traditional class, we rarely talk during the lesson. Teachers do most of the talking. They ask some questions but not much time is spent on the answers, so usually it is hard to get a chance to talk in the class. In this course, we had a lot of time to talk and share our opinions. Hearing others’ opinions also helped us gain a different perspective.*

In their opinion, the chance to work with peers also contributed to the positive classroom environment. As they reported, in group work, PTs felt more comfortable and “did not have the fear of saying wrong things.” One PT explained why she felt less stress in the flipped classroom as follows:

*In a lecture-based course, I easily get lost and miss some information. This creates a lot of stress for me. In this course, I felt relaxed because I did not have to follow the teacher’s lecture during classroom time.*

The third theme emerging from the data was the benefits of individualised learning. In their comments, PTs emphasized the importance of self-pacing. They all appreciated watching the lectures at any time they wanted and being able to decide on their own pacing. They paused, rewound and replayed the video lectures as they needed. Here are PTs’ comments illustrating these points:

*I do not feel ready to learn all the time. I watched the video lectures when I felt ready. I watched the lectures whenever I wanted—sometimes while travelling on the bus, sometimes while playing a game or eating.*

*Before the exam, I watched all the videos again. They really helped me prepare for the exam.*

*While watching the lectures, I replayed the sections that were complicated again and again.*

During the interview, as PTs were discussing how much they benefitted from the flipped instruction, the idea of applying it to their future classes emerged as the fourth theme in the data. Although PTs did not exactly know how to flip a language class, they all believed in its advantages for students’ learning and engagement. They all mentioned the lack of time for practice in language classrooms and stated that in Turkey, language classes are mostly teacher-centred, and students do not get enough opportunities to practice their language skills. Here are the relevant comments of the PTs:

*Students can watch some grammar videos at home and do practice activities in the classroom.*

*In crowded classrooms, students do not always get the chance to talk. This situation demotivated me when I was in high school. The flipped model can be a solution to this problem.*

*The flipped class model is very enjoyable. Students would love it.*

*I will do some research about flipping a language classroom. I believe my future students will benefit from it. Today’s students love learning with technology, so the flipped model will be very engaging for them.*

**Discussion of the findings**

The present study was designed to measure the efficacy of the flipped classroom model as applied to a classroom management course in a pre-service English teacher education program. The flipped classroom was compared to a traditional lecture-based classroom in order to determine its effect on PTs’ perceived self-efficacy beliefs and learning outcomes. Findings revealed that the PTs in the flipped classroom had more developed self-efficacy beliefs and scored higher on the final exam than the PTs in the traditional classroom. The flipped group PTs also felt well prepared and confident about dealing with classroom management issues in their future classes; moreover, their perceptions of the flipped classroom approach were positive.
Tschannen-Moran and Hoy (2001) defined a teacher’s efficacy belief as “a judgment of his/her capabilities to bring about desired outcomes of student engagement and learning, even among those students who may be difficult or unmotivated” (p. 783). The vast majority of research has indicated self-efficacy as one of the important predictors of teachers’ instructional effectiveness and behaviour management (e.g., Gibson & Dembo, 1984; Welch, 1995). In the present study, the flipped group PTs’ self-efficacy beliefs improved more than the beliefs of the control group PTs. The improvement was statistically significant according to the overall scale and the subscales of student engagement and classroom management.

Among the sources influencing teachers’ self-efficacy beliefs are (1) mastery experience, (2) vicarious experience, (3) verbal or social persuasion, and (4) physiological arousal or emotional state (Bandura, 1977). Mastery experience, defined as satisfaction with professional performance, is considered to be the most important determinant of self-efficacy. Bandura (1977) state that “successes raise mastery expectations; repeated failures lower them” (p. 195). PTs in the flipped classroom worked collaboratively to role-play classroom management scenarios. They developed classroom management strategies to deal with given situations and applied them in their demonstration. They also reflected on the strategies employed in terms of their potential effectiveness in a real classroom setting. Acting as a teacher, applying a strategy to a behaviour problem, and evaluating its possible consequences might have provided PTs with authentic evidence of their performance in such classroom situations. Vicarious experience has been identified as the second most potent influence on one’s self-efficacy beliefs. Vicarious learning occurs by observing the behaviours of others (Bandura, 1977). In this study, PTs in the flipped classroom observed their peers’ demonstrations of classroom scenarios and analysed video-recorded lessons particularly focusing on the use of classroom management strategies. Social persuasion in terms of feedback received from others represents another important source of self-efficacy beliefs (Bandura, 1977). PTs in the flipped group received constant feedback from their peers and the instructor on their instructional choices and classroom management strategies, which might have contributed to their self-efficacy development. Finally, physiological states such as stress and anxiety also provide information about self-efficacy beliefs (Bandura, 1977). In the present study, the flipped group PTs emphasized the positive learning environment they were engaged in and stated how relaxed and confident they felt in the classroom. As mentioned in the interview, PTs were not afraid of making a mistake or wrong decision. They shared their opinions comfortably and always received constructive feedback from their peers and the instructor.

Another major finding of the study was about the flipped group PTs’ significantly higher achievement on the final exam. This finding showing an improved learning outcome confirms the findings of several studies that focused on the efficacy of the flipped classroom method on improving students’ academic achievement (e.g., Davies et al., 2013; McLaughlin et al., 2014; Marcey & Brint, 2012). In the present study, PTs stated that they watched the lecture videos any time they wanted and as many times as they needed. They also moved through the content at their own pace. As the ability of pausing, rewinding, fast-forwarding and replaying has been shown to increase students’ engagement (Day & Foley, 2006), the increase in PTs’ learning outcomes might stem from the self-pacing opportunity in the learning environment.

PTs’ comments in the interview revealed an increase in their perceived engagement, which confirmed the findings of McLaughlin et al. (2014), Strayer (2009), and Rowe, Frantz and Bozalek (2013). Similar to the subjects in those studies, PTs in the present study perceived themselves to be more motivated in the flipped classroom compared to the traditional lecture-based classroom. As they stated, their preference for the flipped classroom was due to its allowance for active participation in the classroom. PTs found sitting passively and listening to a lecture boring. Instead, they preferred being actively engaged in the lesson.

In the interview, PTs also discussed flipping their future classrooms considering the benefits they experienced in the flipped course. Learning in such an environment helped them gain first-hand experience of the flipped classroom approach. PTs’ use of technology throughout their teacher education experiences has been found to be the most significant predictor of their technology integration (Brent, Brawner, & Van-Dyk, 2002; Snider, 2002). Therefore, PTs’ involvement in the flipped instruction might contribute to their future practices with technology.

**Conclusion**

The call for reform in higher education due to the advent of new technologies requires changes to traditional pedagogy. The flipped classroom approach allows for such a pedagogical shift to create a student-centred, individualised learning environment based on the constructivist theory of learning. The present study adds to the growing field of literature about the flipped approach. Furthermore, it confirms the findings of similar studies by
presenting evidence for PTs’ improved self-efficacy beliefs, academic achievement and favourable perceptions about the flipped classroom as an innovative instructional approach in a teacher education course.

The flipped course model in this study might inform the instructional design and delivery in higher education institutions by proposing a framework based on the principles of constructivism and differentiated instruction. In- and out-of-class activities used in this study also have the potential to guide instructors in their own designs as the activities considered various aspects of student engagement and learning in the flipped approach. Moreover, this study is unique in terms of the context in which it was embedded. To date, there are no studies clearly depicting the flipping process of a teacher education course in the Turkish context. Therefore, a detailed description of the planning and implementation processes might be informative for the design of similar instructional models in this specific context.

The study has its limitations too. First, the sample size was relatively small with 62 total participants, 32 in the experimental group and 30 in the control group. For this reason, some caution should be taken when generalising the findings. Second, since the instructor of the course was also the researcher, unintended bias may have influenced the outcomes.

To conclude, the flipped classroom might be applied to other teacher education courses as a viable pedagogical model for reversing the design of a classroom to enhance PTs’ learning experiences and their future teaching practices in the 21st century.

References


Baker, J. W. (2000). The “Classroom Flip”: Using web course management tools to become the guide by the side. In J. A. Chambers (Ed.), Selected papers from the 11th International Conference on College Teaching and Learning (pp. 9-17). Jacksonville, FL: Florida Community College at Jacksonville


Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. Eugene, OR: International Society for Technology in Education.


Using “First Principles of Instruction” to Design Secondary School Mathematics Flipped Classroom: The Findings of Two Exploratory Studies

Chung Kwan Lo* and Khe Foon Hew
The University of Hong Kong, Hong Kong // cklohk@hku.hk // kfhew@hku.hk
*Corresponding author

ABSTRACT
Flipping the classroom is a current pedagogical innovation in many schools and universities. Although interest in flipped classroom (or Inverted Classroom) continues to grow, its implementation so far has been driven more by teachers’ intuitive beliefs, rather than empirically-based principles. Many studies merely replace in-class instructions with videos and use class time for group discussions. But what instructional design framework should we use in planning the overall flipped classroom approach? This paper attempts to answer this question through two exploratory studies conducted in a Hong Kong secondary school. In Study 1, a flipped classroom Mathematics remedial approach was offered for underperforming students (n = 13) in Form 6 (Grade 12). In Study 2, high ability students (n = 24) in Form 6 participated in another flipped classroom Mathematics training approach. Both flipped classroom approaches utilized the First Principles of Instruction design theory. Paired t-test results indicated significant learning gains in both groups of students. Based on the suggestions of students and teacher as well as the existing literature, several recommendations for course planning, out-of-class learning, and in-class learning of flipped classroom are proposed.

Keywords
Flipped classroom, Inverted classroom, First principles of instruction, Mathematics, Pedagogy

Introduction
Flipped Classroom is a technology-supported pedagogical innovation which has become popular in recent years. According to Bishop and Verleger (2013), flipped classroom consists of two components: (1) Direct computer-based individual instruction outside the classroom, and (2) Interactive group learning activities inside the classroom. In the out-of-class learning component, students watch instructional videos prepared by teachers. Students thus acquire some basic information before the face-to-face lesson. The in-class time is then freed up for more interactive learning activities such as collaborative problem solving and receiving teacher’s individual assistance.

Hamdan, McKnight, McKnight, and Arfstrom (2013) argue that flipped classroom is a feasible strategy which caters to the needs of diverse learners. For example, if students do not understand the materials presented in the video lectures, they can pause or replay the instruction videos for revision. At the same time, high ability students can skip certain parts of the video lectures to save their learning time. As for the face-to-face lessons, since the in-class time is no longer occupied by direct teaching, more time can be spent on the teacher’s one-to-one assistance and small-group tutoring for the less capable students (Bergmann & Sams, 2009), or problem-based learning and small-group learning activities which are suitable for high ability students (Matthews & Dai, 2014). However, Hamdan et al. (2013) lament that there is a lack of empirical study that investigates the use of flipped classroom for diverse learners. In fact, most of the existing studies of flipped classroom focused on flipping a particular course (see Bishop & Verleger, 2013; Giannakos, Krogstie, & Chrisochoides, 2014; O’Flaherty & Phillips, 2015 for a review) rather than explicitly examining whether flipped classroom can benefit underperforming or high ability students.

Besides the lack of studies that examines how flipped classroom may help diverse students, there are two other limitations of previous flipped classroom research. First, a majority of studies has been conducted in Western higher education sector (see Bishop & Verleger, 2013; Giannakos et al., 2014; O’Flaherty & Phillips, 2015 for a review). Very few published studies have hitherto focused on the Asian secondary school settings. Contrary to the popularity in the West, Subramaniam (2008) suggested that contemporary education approaches such as online education may not necessarily capture Asian learners’ interest and engagement. Some Chinese learners’ preference for teacher-centered learning, and classroom learning may adversely affect the efficacy of flipped classroom. In a traditional class, students typically learn about the subject matter through a teacher-led lecture, followed by a teacher-led activity during class time. However, students in a flipped class are required to take more responsibility for their own learning such as watching the video lectures before class, and participating in group problem-solving activities during in-class lessons. Some Asian secondary school students, being typically passive during in-class sessions, barely interacted with other students; they merely sat quietly and waited for the teacher to approach them (Nawi et al., 2015). This therefore raises several intriguing questions: How would
students in a Hong Kong secondary school perceive the use of flipped classroom? Would they find the flipped classroom approach more engaging than the traditional classroom instruction method? Our present study aims to address these very questions.

Second, many studies discussed what benefits can be expected from flipping the class (e.g., Fulton, 2012; Gannod, Burge, & Helmick, 2008), but fell short of defining and examining the design principles of flipped classroom (Kim, Kim, Khera, & Getman, 2014) or utilizing a conceptual framework that could guide the design of flipped classroom (O’Flaherty & Phillips, 2015). Currently, the design of flipped classroom has often been limited to the practice of merely replacing in-class instruction with video-recorded lectures and using class time for homework (Kim et al., 2014). But what instructional design framework should we use in planning the overall flipped classroom approach?

The present study aims to extend our collective understanding of flipped classroom in three ways. First, we tested the feasibility of using an instructional design theory – Merrill’s (2002) First Principles of Instruction to implement flipped classroom. The effectiveness of the “First Principles of Instruction” had been examined in a study undertaken by Thompson/Netg, a company that offers learning solutions for individuals, businesses and institutions (Thomson, 2002). Using a three-group experimental design, the investigators found that the group which received instruction developed based on the “First Principles” scored the highest scores than the other two groups. All differences were statistically significant. Further, the “First Principles” group managed to complete three course activities in the shortest time (29 minutes), compared to the group that received the existing commercial version of the company’s course (49 minutes), while most of the control group failed to finish the tasks. Studies done by other researchers (e.g., Frick, Chadha, Watson, & Zlatkovska, 2010) have also suggested that the use of First Principles of Instruction can improve students’ motivation and learning when compared with other forms of instruction. The First Principles of Instruction design theory therefore provides us with a unique theoretical framework to implement our flipped classroom approach. Second, we extended our study to a Hong Kong secondary school context; more specifically to the teaching and learning of Form 6 (Grade 12) Mathematics. Third, we designed and offered two flipped classroom for underperforming students and high ability students correspondingly. The effectiveness, student perceptions, and teacher’s experiences of the two Flipped Classrooms could thus be compared.

Two exploratory studies were conducted: Study 1 investigated a flipped classroom remedial approach for underperforming students, and Study 2 examined the effects of a flipped classroom approach for high ability students. The following research questions were addressed:

- To what extent does the use of flipped classroom have an impact on underperforming and high ability students’ Mathematics learning?
- How do the teacher and students perceive the use of flipped classroom?
- How can the design and implementation of flipped classroom be improved?

**Flipped classroom design**

The two studies reported in this paper were distinct in terms of student cohorts but taught by the same teacher. Study 1 was designed for underperforming students while Study 2 was for high ability students. Despite the different student cohorts, both Studies shared certain similarities in terms of the overall design of the flipped classroom approach, the data sources, and the statistical analyses used. Each flipped classroom approach was designed based on Merrill’s (2002) First Principles of Instruction design theory. Malone (1985) explains that unlike explanatory theory (“Y because of X”), design theories emphasize how to achieve goals (“In order to achieve Y, do X”). The First Principles of Instruction (see Figure 1 and Table 1) are largely context-free, and are derived from a review of several instructional design theories and models such as the Vanderbilt Learning Technology Center’s Star Legacy (Schwartz, Lin, Brophy, & Bransford, 1999), Constructivist Learning Environment model (Jonassen, 1999), the Four Component Instructional Design model (van Merriënboer, 1997), and Learning by Doing model (Schank, Berman, & Macperson, 1999).

![Figure 1. Merrill’s (2002) First Principles of Instruction](attachment:image.png)
Table 1. First Principles of Instruction (summarized from Merrill, 2002, p. 45-50)

<table>
<thead>
<tr>
<th>Instructional principles</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning is promoted when learners are engaged in solving problems that can be found in the real world [Problem-centric]</td>
<td>Show task: Learning is promoted when learners are shown the task that they will be able to do or the problem they will be able to solve as a result of completing a module or course.</td>
</tr>
<tr>
<td>Learning is promoted when existing knowledge is activated as a foundation for new knowledge [Activation]</td>
<td>Previous experience: Learning is promoted when learners are directed to recall, or relate knowledge from relevant past experience that can be used as a foundation for the new knowledge.</td>
</tr>
<tr>
<td>Learning is promoted when new knowledge is demonstrated to the learner [Demonstration]</td>
<td>Demonstration: Learning is promoted when the teacher demonstrates the appropriate procedures to solve the problems.</td>
</tr>
<tr>
<td>Learning is promoted when new knowledge is applied by the learner [Application]</td>
<td>Practice: Learning is promoted when the activities and the tests are consistent with the stated learning objectives. Varied problems: Learning is promoted when learners are required to solve a set of varied problems.</td>
</tr>
<tr>
<td>Learning is promoted when new knowledge is integrated into the learner’s world [Integration]</td>
<td>Creation: Learning is promoted when learners can use their new knowledge or skill to solve more advanced problems.</td>
</tr>
</tbody>
</table>

Figure 2. Overarching design framework of flipped classroom

The First Principles of Instruction design theory provides us with a unique theoretical framework to implement our flipped classroom approach (see Figure 2). Specifically, in our flipped classroom approach used in both Studies, we delivered the activation phase, demonstration phase, and application phase outside classroom via video lectures. Students first watched several instructional videos, as mini-lectures, for a particular topic (e.g., mid-point formula in coordinate geometry) at home. In each mini-lecture, the teacher would first show the task that students were able to handle after the completion of the mini-lecture [problem-centric – show task]. The teacher then activated students’ prior knowledge by recalling relevant concepts or knowledge previously learned [activation phase]. Next the teacher demonstrated the new knowledge, strategy, or procedure for solving the problem [demonstration phase]. The mini-lectures could be paused at any time or be played back repeatedly so that students could learn at their own pace. After viewing the mini-lectures, students would answer some simple online quizzes by applying what they had learned in the video lecture to promote learning [application phase]. The online quizzes helped teachers check the students’ learning by analyzing their responses to the questions. During face-to-face class sessions, we delivered the activation phase, application phase, and integration phase inside the classroom. The teacher would first review the topics covered in the video lecture, and clarify any
misunderstandings [activation phase]. Students would then apply the concepts learned in solving some simple problems either individually or in pairs [application phase]. Students were also asked to apply their knowledge in solving more advanced or real-world problems in groups under the supports of teacher and peers [integration phase]. The use of group discussion could deepen students’ understanding and help them integrate the new knowledge into real-world contexts (Warter-Perez & Dong, 2012). Figure 3 shows the flow of teaching and learning activities in each session.

![Figure 3. The flow of teaching and learning activities in each session](image)

**The design of instructional videos**

The design of our video lectures was informed by evidence-based findings. First, we limited the length of our instructional videos to less than six minutes. Videos shorter than six minutes were found to be most engaging to students (Guo, Kim, & Rubin, 2014). Second, we followed the guidelines pertaining to the cognitive theory of multimedia learning (Mayer, 2014). For example, it is suggested that learning is enhanced when extraneous material are excluded (i.e., coherence effect), when cues are provided to highlight essential materials (i.e., signaling effect), and when words are spoken in conversational style (uses I and you as in an informal conversation with the learner) rather than non-personalized style in which the teacher speaks in a third-person formal monologue (i.e., personalized effect) (Mayer, 2014). Third, we mainly employed Khan-style tutorial style (i.e., a teacher drawing on a digital tablet) since the natural motion of human handwriting can be more engaging than static computer-generated fonts (Cross, Bayyapunedi, Cutrell, Agarwal, & Thies, 2013). Figure 4 shows a screen-shot of a video lecture used in Study 1.

![Figure 4. Screen-shot of the digital tablet drawing lecture in Study 1](image)
General method

Overview of the two studies

The two exploratory studies were conducted in a Hong Kong secondary school. Most of the students have minimal experience of using flipped classroom. Table 2 summarizes the contexts of Study 1 and Study 2.

<table>
<thead>
<tr>
<th>Number of participant</th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Form 6 (Grade 12)</td>
<td>Form 6 (Grade 12)</td>
</tr>
<tr>
<td>Mathematics ability of participants</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Course topic</td>
<td>Coordinate geometry</td>
<td>Arithmetic and geometric sequences and their summations</td>
</tr>
<tr>
<td>Number of session</td>
<td>Three</td>
<td>Six</td>
</tr>
<tr>
<td>Length of each instructional video</td>
<td>≤ 6 minutes (two to three videos per session)</td>
<td>≤ 6 minutes (two to three videos per session)</td>
</tr>
<tr>
<td>Time for each face-to-face lesson</td>
<td>50 minutes</td>
<td>One hour</td>
</tr>
<tr>
<td>Duration of program</td>
<td>Two weeks</td>
<td>Four weeks</td>
</tr>
</tbody>
</table>

Data collection and analysis

The three research questions were addressed by using four major sources of data, including pre-test and post-test, questionnaire survey (for Study 2 only), student interview, and teacher interview. Table 3 shows each research question and the methods associated.

<table>
<thead>
<tr>
<th>Research question</th>
<th>Data source</th>
<th>Data analysis method</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1: To what extent does the use of flipped classroom have an impact on underperforming and high ability students’ Mathematics learning?</td>
<td>Pre-test and post-test</td>
<td>Paired sample t-test</td>
</tr>
<tr>
<td></td>
<td>Questionnaire survey</td>
<td>Descriptive statistics</td>
</tr>
<tr>
<td>RQ2: How do the teacher and students perceive the use of flipped classroom?</td>
<td>Student interview</td>
<td>Coded and organized into emerging categories</td>
</tr>
<tr>
<td></td>
<td>Teacher interview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Questionnaire survey</td>
<td></td>
</tr>
<tr>
<td>RQ3: How can the design and implementation of flipped classroom be improved?</td>
<td>Student interview</td>
<td>Coded and organized into emerging categories</td>
</tr>
<tr>
<td></td>
<td>Teacher interview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Questionnaire survey</td>
<td></td>
</tr>
</tbody>
</table>

To answer RQ1, a 15-minute pre-test and 15-minute post-test were conducted to assess students’ learning progress. To enhance the reliability and validity, all test questions were adopted and modified from the Mathematics public examinations in Hong Kong. By referring to the annual reports of the public examination, we ensured that the questions in pre-test and post-test were different but similar in terms of scope and difficulty level. To evaluate the effectiveness of the intervention, paired sample t-test was used to compare the difference between pre-test mean and post-test mean. Besides the use of tests scores, questionnaire data could also reveal the general impact on student learning.

To answer RQ2 and RQ3, a 15-minute questionnaire and student interview were used to study student perception of flipped classroom. The questionnaire was adopted and modified from Johnson’s (2013) survey. The questionnaire survey asked students to rate their general attitude toward the flipped classroom designed. Additional spaces were provided for free text responses. In the student interview, we investigated how students learn through flipped classroom, examined their perceptions and experience, and identified any difficulties encountered. An interview protocol of suggested questions and possible follow-up questions was designed based on Lofland, Snow, Anderson, and Lofland’s (2006) guideline on interview. In particular, some of the interview questions were adopted from Zappe, Leicht, Messner, Litzinger, and Lee’s (2009) survey of flipped classroom.

During the teacher interview, the teacher was asked to reflect upon his implementation of flipped classroom according to a guiding protocol. The protocol focused on two areas: (1) The perceptions of implementing flipped classroom; and (2) the difficulties encountered in flipped classroom.
Quantitative data from questionnaires provided a general understanding of students’ perceptions of flipped classroom. The qualitative data collected from the questionnaires and interviews were thematically analyzed and organized into categories (Corbin & Strauss, 2008).

**Study 1: Flipped classroom for underperforming students**

**Participants and procedure**

Participants were 13 Form 6 (Grade 12) students. They were invited because of their underachieving performance in coordinate geometry. The remedial program was thus about coordinate geometry. Referring to the curriculum guides of Hong Kong Secondary Mathematics education (CDC & HKEAA, 2014), the class schedule of the remedial program was set (see Table 4). While the video lectures demonstrated some basic information of the topic (e.g., calculating the distance between two points), the in-class time mainly focused on handling the more advanced problems and real-world problems. Figure 5 shows one of the real-world problems used in the program.

<table>
<thead>
<tr>
<th>Session</th>
<th>Video lecture (out-of-class)</th>
<th>Face-to-face lesson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mid-point of two points; Distance between two points; and Slope of straight line</td>
<td>Transformation of point Advanced problems</td>
</tr>
<tr>
<td>2</td>
<td>Equation of straight line; x- and y-intercept of straight line; and Interception point of straight lines</td>
<td>Perpendicular lines Advanced problems</td>
</tr>
<tr>
<td>3</td>
<td>Slope of the equation of straight line; Line perpendicular to straight line; and Perpendicular bisector of two points</td>
<td>Concept of locus Real-world problems</td>
</tr>
</tbody>
</table>

![Sample question of the real-world problems of coordinate geometry](image)

*Figure 5. Sample question of the real-world problems of coordinate geometry (see Chik, 2014)*

**Results**

**Pre-test and post-test**

The total score of pre-test and post-test was both 10. The questions in the two tests were different but similar in terms of scope and difficulty level. A paired t-test showed a significant difference between the pre-test mean \((n = 13, M = 2.77, SD = 1.79)\) and the post-test mean \((n = 13, M = 5.85, SD = 2.41)\), \(t(12) = 6.50, p < .0001\). The Cohen’s \(d\) value was 1.80, indicating a large effect size. Figure 6 shows the box plot and the results of the pre-test and post-test scores.

![Box plot and summary of pre-test and post-test scores](image)

*Figure 6. Box plot and the summary of the pre-test and post-test results of the remedial program in Study 1*
Students’ perceptions of flipped classroom

Student interview data was thematically analyzed and organized into several categories, namely course content and design, collaboration with peers, and teacher’s supports. Some direct quotations from the interview findings are reported for illustration.

Course content and design. Almost all students supported the use of flipped classroom and perceived that this instructional approach facilitated their learning. Most of the students reported that they could review the materials anytime and anywhere: “We can review the videos when necessary” (Student 1). However, a few students reported that they could not handle some of the advanced problems and real-world problems. They requested more basic exercises to help them acquire the new knowledge and skills: “The final problem (real-life problem) is very difficult. … I need to do more exercises. In this way, I can master the skills better” (Student 2).

Collaboration with peers. Almost all students engaged in small-group learning activities. A few students explained that they could support each other by discussing problems, explaining concepts, and checking answers or steps of problem solving: “I find learning in groups better since my classmate can answer my questions immediately when I don’t understand” (Student 7).

Teacher’s supports. The most commonly mentioned issue concerned the support of out-of-class learning. A number of students expressed that they could not receive help during the video lectures. Some of them suggested the teacher provide explanations or solutions to the online exercises: “we cannot ask question immediately while watching videos” (Student 5); “Please provide a full solution and explanation of the online exercises, especially the harder one” (Student 8).

Study 2: Flipped classroom for high ability students

Participants and procedure

There were 117 students in the Form 6 non-science classes. Based on their latest Mathematics examination score, the top 25% of the students were invited on a voluntary basis. A total of 24 students participated in this training program. The course was about arithmetic and geometric sequences and their summations. Referring to the curriculum guides of Hong Kong Secondary Mathematics education (CDC & HKEAA, 2014), the class schedule of the training program was set (Table 5). The video lectures handled the basic parts of the topic (e.g., evaluating the summation of an arithmetic sequence). As for the face-to-face lessons, some advanced application problems and real-world problems were discussed, such as counting the number of seats in a theatre (a problem of arithmetic sequences), and calculating the amount of revenue of a firm (a problem of geometric sequences).

Table 5. Overview of the class schedule of the training program in Study 2

<table>
<thead>
<tr>
<th>Session</th>
<th>Video lecture (out-of-class)</th>
<th>Face-to-face lesson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Review on sequences; and Introduction to arithmetic sequences</td>
<td>Advanced problems</td>
</tr>
<tr>
<td>2</td>
<td>Introduction to geometric sequences</td>
<td>Advanced problems</td>
</tr>
<tr>
<td>3</td>
<td>Distinguishing between arithmetic sequences and geometric sequences; and Introduction to summation of sequence</td>
<td>Real-world problems</td>
</tr>
<tr>
<td>4</td>
<td>Summation of an arithmetic sequence</td>
<td>Real-world problems</td>
</tr>
<tr>
<td>5</td>
<td>Sum of the first n terms of a geometric sequence</td>
<td>Real-world problems</td>
</tr>
<tr>
<td>6</td>
<td>Sum to infinity of a geometric sequence</td>
<td>Real-world problems</td>
</tr>
</tbody>
</table>

Results

Pre-post test

The total score of pre-test and post-test was both 15. The questions in the two tests were different but similar in terms of scope and difficulty level. A paired t-test showed a significant difference between the pre-test mean \( (n = 24, M = 2.00, SD = 1.77) \) and the post-test mean \( (n = 24, M = 8.08, SD = 3.03), t(23) = 9.43, p < .0001 \). The Cohen’s \( d \) value was 1.92, indicating a large effect size. Figure 7 shows the box plot and the results of the pre-test and post-test scores.
Figure 7. Box plot and the summary of the pre-test and post-test results of the training program in Study 2

Students’ perceptions of flipped classroom

Table 6 shows the questionnaire results of Study 2. Overall, most of the students (87.5%) found that flipped classroom was more engaging than traditional classroom, and preferred learning at their own pace. Also, many students (70.8%) liked watching instructional videos, and recognized that flipped classroom provided more chances for peer communication.

<table>
<thead>
<tr>
<th>Item</th>
<th>Score on the 5-point Likert Scale (% Respondents)</th>
<th>Mean(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 The flipped classroom is more engaging than traditional classroom instruction</td>
<td>54.2 33.3 12.5 0 0</td>
<td>4.42 (.72)</td>
</tr>
<tr>
<td>2 I like watching the lessons on video</td>
<td>41.7 29.2 29.2 0 0</td>
<td>4.13 (.85)</td>
</tr>
<tr>
<td>3 I prefer a video-recording of the lesson to a traditional teacher-led lesson</td>
<td>45.8 25.0 29.2 0 0</td>
<td>4.17 (.87)</td>
</tr>
<tr>
<td>4 I like to self-pace myself through the course</td>
<td>50.0 37.5 12.5 0 0</td>
<td>4.38 (.71)</td>
</tr>
<tr>
<td>5 I like taking my quizzes online by using online learning platform</td>
<td>45.8 20.8 29.2 4.2 0</td>
<td>4.08 (.97)</td>
</tr>
<tr>
<td>6 The flipped classroom gives me more chances to communicate with other students.</td>
<td>33.3 37.5 29.2 0 0</td>
<td>4.04 (.81)</td>
</tr>
<tr>
<td>7 I am more motivated to learn in the Flipped Classroom</td>
<td>41.7 37.5 20.8 0 0</td>
<td>4.21 (.78)</td>
</tr>
<tr>
<td>8 The flipped classroom has improved my learning of Mathematics</td>
<td>41.7 45.8 12.5 0 0</td>
<td>4.29 (.69)</td>
</tr>
</tbody>
</table>

Note. 5 = strongly agree to 1 = strongly disagree.

Similar to Study 1, students’ responses to the open-ended questions in the questionnaire, and interviews were thematically analyzed and organized into several categories, namely course content and design, collaboration with peers, and teacher’s supports.

Course content and design. Students talked about the advantages of flipped classroom, such as being able to learn at their own pace, and having autonomy in learning: “Students are free to choose whether to watch the videos for revision or not” (Student 10), “We can decide our own learning progress” (Student 9). In addition, students’ perceptions of flipped classroom were generally positive. Some students even requested the teacher to provide more examples and exercises for them, as well as extend the duration of lessons: “It would be better to provide more examples and advanced application problems” (Student 15); “We can stay even after 5:30pm (end of the lesson)” (Student 20).

Collaboration with peers. Most students stated that the in-class discussion facilitated their learning. They also valued the communication with peers in their learning. For example, “Students mainly discussed the solution in class, which facilitated our communication and learning” (Student 18).

Teacher’s supports. While many students appreciated that they could receive more help from teacher during in-class time, a number of students expressed that they could not get immediate assistance in their out-of-class learning. A few students asked for a place for posting questions to the teacher: “We cannot get instant feedback when we encounter problems at home” (Student 17); “You can provide a place for students to ask questions during the out-of-class session” (Student 12).
Teacher’s overall opinions of the flipped classroom

The teacher interview data were thematically analyzed and organized into two main categories, namely his experiences of implementing flipped classroom, and the difficulties encountered during its implementation.

Experiences of flipped classroom. The First Principles of Instruction provided a clear guideline for the teacher to design flipped classroom, instead of merely relying on his intuitive beliefs. To facilitate the learning of new content, the teacher affirmed that recalling the relevant pre-requisite knowledge was necessary. By putting some revision videos online, students could do the revision outside the classroom, and more in-class time was thus spent on clarifying misunderstanding or solving more advanced problems. However, in addition to activation outside the classroom, in-class activation was useful for both underperforming and high ability students. This is because the teacher observed that some students might forget what they had learned out-of-class when coming back to the classroom: “They could not recall the knowledge because they had visited the video lecture too early before its corresponding lesson.”

While demonstrating the new content via video lectures could free-up more in-class time, the teacher thought that parts of the course were still suitable to deliver inside the classroom, especially for the difficult learning items. This was because the teacher found it difficult to explain the complicated concepts in a short video. Also, teacher could have a better understanding of whether students could follow the presentation in a face-to-face teaching and learning environment: “Students’ facial expressions usually give me some hints, telling me which parts are difficult for them and I need to explain further.”

Difficulties encountered. The teacher found that the analytics of the online quizzes were useful in lesson preparation. However, he was concerned that some students might complete the online quizzes casually: “Based on the results of the online quizzes, I can figure out whether I need to re-teach or not. But I am not sure whether the students completed these multiple choice questions seriously or not.” The teacher also pointed out the importance of engaging in the advanced and real-world problems in Hong Kong secondary education context. At the same time, he admitted that not all students could handle these problems due to their ability and subject interests. He recommended exploring further strategies in flipped classroom to cater to the needs of different students: “In the context of Hong Kong Mathematics education, it is necessary to equip students to solve real-world problems. In public examination, it is not unusual that some questions are related to everyday life. Although the flipped classroom approach provides room for us to handle these problems inside the classroom, it is difficult to satisfy all students due to the time constraint and the large class size.”

General discussion

Contrary to many previous published studies, the present study is distinctive in the following two ways. First, it tested the feasibility of using the Merrill’s (2002) First Principles of Instruction design theory to implement flipped classroom in a secondary school context. A majority of previous studies did not explicate any specific conceptual framework to help instructors design their flipped classrooms (Bishop & Verleger, 2013; Giannakos et al., 2014). Second, very few previous studies utilized their results to develop design principles for using flipped classroom (O’Flaherty & Phillips, 2015). Our present study proposed several recommendations (Table 7) based on the suggestions of students and teacher, as well as relevant literature. The results are discussed in three main sections: Impact on students’ Mathematics learning, the First Principles of Instruction, and a comparison of the two flipped classrooms.

Impact on students’ Mathematics learning

From the pre-test and post-test results, there was a significant learning gain in both Studies. Moreover, the effect size of both Studies was large. From the student interviews of Study 1, most of the students reported that flipped classroom facilitated their learning, which confirmed the test results. In Study 2, the test results were also consistent with their self-perceived learning. There were 87.5% of students who agreed or strongly agreed that “The flipped classroom has improved my learning of Mathematics.” Also, 79.2% of students agreed or strongly agreed that “I am more motivated to learn in the flipped classroom.”

Nevertheless, we do not claim that flipped classroom is better than other instructional approaches in other contexts. In the present study, we could only suggest that the use of flipped classroom may be useful in
increasing the Mathematics achievement for both underperforming and high ability students. We could not claim causality effect since no control group (e.g., non-flipped classroom condition) was employed.

**Merrill's (2002) First Principles of Instruction**

In the following parts, we discuss each of the four phases of First Principles of Instruction, namely activation, demonstration, application, and integration.

**Activation**

The teacher shared that he “would usually have a revision on the background information before teaching some new materials” even in traditional classroom, aiming to “prepare their learning by recalling the relevant knowledge.” But in his experience, not every student was engaged in this revision, especially for the high ability students. He explained that “They already have the knowledge in their mind. In-depth revision was not necessary for them but for the less capable students.” Therefore he affirmed that it is desirable to shift the revision part outside the classroom. In students’ opinion, they can benefit since “Students are free to choose whether to watch the videos for revision or not” and “can review the videos when necessary.”

However, we still suggest including a brief review at the beginning of each face-to-face lesson. Based on the teacher’s observation, some of the students forgot what they had learned in the video lectures when coming back to the class. But looking at their performances of the online exercises, the teacher believed that these students had prepared for the class seriously. He argued that “They could not recall the knowledge because they had visited the video lecture too early before its corresponding lesson.” In this regard, in-class activation on out-of-class learning materials may be useful for both underperforming and high ability students. As Munson and Pierce (2015) recommended, a brief review highlighting the key concepts presented in the video lecture can serve as a starter of class.

**Demonstration**

In the teacher’s opinion, “direct demonstration is still an effective way to deliver new concepts for my students.” For the simple learning items, he found that teaching via instructional videos was similar to the direct teaching inside the classroom. But the advantage of using instructional videos is to free-up the in-class time for interactive group learning activities (Bishop & Verleger, 2013). However, a number of students expressed that they could not ask questions immediately and get instant feedback when watching the instructional videos or doing the online quizzes. Therefore, we suggest creating a Q & A forum or allowing students to leave comments in the online learning platform. In this way, both teachers and their classmates can provide timely feedback when students post their questions online. Conte et al. (2015) further recommended enabling “real-time question-and-answer interactions and a full archive of all information exchanged” (p. 70).

**Application**

Online quizzes are useful for students to apply the knowledge. There were 66.7% of the students in Study 2 agreed or strongly agreed that “I like taking my quizzes online by using online learning platform.” The analytics of the online quizzes also provided information for teaching preparation (Mok, 2014). He mentioned that “Based on the results of the online quizzes, I can figure out whether I need to re-teach or not.” But at the same time, he was “not sure whether the students completed these multiple choice questions seriously or not. It is possible that students submitted their quizzes by randomly clicking an answer.” Therefore, he had designed a set of pre-lesson worksheets which required students to write down some content notes of the instructional videos, and display some problem solving steps of several simple exercises. Similar to Clark (2015) and Little (2015), the teacher could assess students’ pre-class preparation by checking their worksheets.

In his experience, the teacher noticed that “There is a gap between understanding and applying the concepts.” In traditional classroom, students cannot get help immediately from their teacher or peers when doing their homework. In the present study, the homework problems were handled inside the classroom in a small-group learning environment. Similar to Clark’s (2015) observation, students in the two Studies “learned from each
other by discussing problems, explaining procedures, and confirming answers” (p. 109). Indeed, a number of students pointed out that “The advantage of flipped classroom is having more discussion” inside the classroom.

When solving a set of varied problems during the application phase of flipped classroom, peer-supported learning is especially important because of the highly interactive nature of this instructional approach. On one hand, teachers are able to spend more in-class time for one-to-one assistance and small-group tutoring (Bergmann & Sams, 2009). But on the other hand, teachers are unavailable to help other students when they are occupied by some students in need. Enfield (2013) lamented that “this resulted in students waiting for long periods of time for help” (p. 26). In this regard, the teacher highly encouraged his students to provide feedback for their classmates:

“Suppose I have answered group 1’s question. When group 2 asks the same question, I would direct group 2 to ask group 1. If I am not occupied by other groups, I would listen to group 1 and see how they explain to group 2. In this way, not only group 2 can get help, but also I can check for group 1’s actual understanding on the question and clarify their concepts when necessary. Sometime I perceived that their wordings were more understandable among them.”

In fact, providing feedback is cognitively engaging (Nicol, Thomson, & Breslin, 2014). Although peers’ feedback may be regarded as lacking expertise, Love, Hodge, Grandgenett, and Swift (2014) found in their flipped classroom study that “explaining a problem or idea to their partner helped them to develop a deeper understanding of it” (p. 322). In our present study, a student affirmed that “Learning in groups is better since my classmate can answer my questions immediately when I don’t understand.” It is thus important for flipped classroom practitioners to develop a routine of peer collaboration (Enfield, 2013).

Integration

To promote student learning, students have to engage in solving more advanced problems and real-world problems (Merrill, 2002). We found that most students were willing to do more advanced application problems. In the words of a student, “The questions were very practical. I perceive that I have learned more when comparing with normal lessons.” However, the teacher realized that not all students were able to handle these kinds of problems, particularly the underperforming students. So how can we address the needs of various students?

We provide an example of catering to diverse learners in flipped classroom. Clark (2015) reported a study of a Secondary School Mathematics Flipped Classroom in the United States. Inside the classroom, students had more time to handle various problems with the supports of their teacher and peers. He further illustrated that the teacher would allow the students to join one of the following three main groups (p. 103-104):

- **Group 1:** Students immediately began working on their independent practice problems without the teacher’s assistance;
- **Group 2:** Students gathered around and reviewed the content with the teacher; and
- **Group 3:** Students congregated at the back of the room and revisited the instructional videos collaboratively on electronic devices.

In this practice, students were free to choose their learning activities. They can join Group 1 if they are able to handle the advanced problems, Group 2 if they need teacher’s assistance and then do more basic exercises to consolidate their knowledge and skills, or Group 3 if they need to re-study with the help of teacher and peers to acquire the out-of-class learning materials. Perhaps each group can be further divided into several sub-groups, and teachers can allow their students to form their own group. According to Self-determination Theory, this practice can satisfy students’ need of autonomy which in turn promotes their intrinsic motivation of learning (Deci & Ryan, 2002). We suggest that teachers should first identify the core materials to be completed by all students. Then different levels of tasks and extra exercises should be prepared for each group correspondingly.

Comparison of the two flipped classrooms

Most of the findings of Study 1 and Study 2 were similar. For student learning, both underperforming students and high ability students achieved a learning gain with a large effect size using the “First Principles of Instruction” enabled flipped classroom approach. Their perceptions of flipped classroom were also positive. Considering the flow of teaching and learning (Figure 3), students in both Studies pointed out the needs of additional supports (e.g., Q & A forum) in the video lecture. For the in-class learning segment, a brief review was recommended since students, regardless of their ability, might forget what they have learned in the video
lecture. Both underperforming students and high ability students recognized that flipped classroom could provide them with a greater chance for peer collaboration. For example, “Students mainly discussed the solution in class, which facilitated our communication.” The teacher’s and peers’ supports outside and inside the classroom are vitally important to support both underperforming students and high ability students in flipped classroom.

**Table 7. Summary of the recommendations of the design and implementation of flipped classroom**

<table>
<thead>
<tr>
<th>Component</th>
<th>Recommendations</th>
<th>Supporting resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course planning</td>
<td>• Identify and prepare the learning materials for the core part of the course (completed by all students), advanced problems (for the high ability students) and extra basic exercises (for underperforming students)</td>
<td>Teacher’s recommendation; students’ recommendation; Clark (2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Out-of-class learning</td>
<td>• Address the activation, demonstration, and application phases</td>
<td>Teacher’s recommendation</td>
</tr>
<tr>
<td></td>
<td>• Limit the duration of instructional videos within six minutes</td>
<td>Guo et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>• Provide revision videos to recall the relevant knowledge for learning new knowledge (especially for underperforming students)</td>
<td>Teacher’s recommendation; Merrill (2002)</td>
</tr>
<tr>
<td></td>
<td>• Enable online question-and-answer interactions with the teacher for students to ask questions and receive immediate feedback</td>
<td>Students’ recommendation; Conte et al. (2015)</td>
</tr>
<tr>
<td></td>
<td>• Provide pre-lesson worksheets to ensure students’ preparation for the class</td>
<td>Teacher’s recommendation; Clark (2015); Little (2015)</td>
</tr>
<tr>
<td></td>
<td>• Prepare the face-to-face lesson based on students’ performances of the online quizzes</td>
<td>Teacher’s recommendation; Mok (2014)</td>
</tr>
<tr>
<td>In-class learning</td>
<td>• Address the activation, application, and integration phases</td>
<td>Teacher’s recommendation</td>
</tr>
<tr>
<td></td>
<td>• Provide a brief review to highlight the key concepts presented in the video lecture to activate students’ prior knowledge</td>
<td>Teacher’s recommendation; Munson and Pierce (2015)</td>
</tr>
<tr>
<td></td>
<td>• Facilitate peer-supported learning by teacher’s encouragement or guideline for students</td>
<td>Clark (2015); Enfield (2013); Love et al. (2014)</td>
</tr>
<tr>
<td></td>
<td>• Design different levels of problem-solving tasks for students (provide more basic exercises for underperforming students and more advanced problems for high ability students)</td>
<td>Students’ recommendation; Clark (2015)</td>
</tr>
<tr>
<td></td>
<td>• Allow students to choose the various learning activities based on their needs (to cater to underperforming students and high ability students)</td>
<td>Clark (2015); Deci and Ryan (2002)</td>
</tr>
</tbody>
</table>

However, students’ views on the in-class problem solving activities were different among the underperforming students and the high ability students. In Study 1, the underperforming students wanted to have more basic exercises because they did not feel confident in doing the advanced problems. For example, “The final problem (real-life problem) is very difficult. … I need to do more exercises. In this way, I can master the skills better.” Through the lens of Merrill’s (2002) First Principles of Instruction, more emphasis should be placed on the application phase before advancing to the integration phase when designing a flipped classroom for underperforming students. In contrast, the high ability students in Study 2 would like to have more advanced and real-world exercises. As suggested by the following student, “It would be better to provide more examples and advanced application problems.” In order to engage more advanced problems, they even asked for extending the class time: “We can stay even after 5:30pm.” Therefore, flipped classroom practitioners can put emphasis on the integration phase to satisfy the needs of high ability students.

As Niemiec and Ryan (2009) suggested, teachers should provide the learning activities which are suitable and optimally challenging for students. Therefore, if the general ability of students is low, teachers should prepare extra basic exercises to consolidate their learning before approaching the advanced problems in their flipped classroom. As for the high ability classes, teachers can provide more advanced and real-world problems for students after dealing with several warm-up exercises. In other words, flipped classroom is not a one-size-fits-all solution for catering to diverse learners. The difficulty and amount of learning materials provided in flipped classroom should match with the ability and needs of students.
Finally, we organize in Table 7 the recommendations discussed above into three main themes, namely course planning, out-of-class learning, and in-class learning. These recommendations are proposed based on the students’ and teacher’s suggestions for improvement from the interviews and open-ended responses of questionnaire. Some of the recommendations are also derived from relevant existing literature. Practically, these recommendations may help practitioners design and implement a flipped classroom.

Conclusion

The flipped classroom approach has become very popular in many educational institutes around the world. In this study, we investigated the use of Merrill’s (2002) First Principles of Instruction design theory to design flipped classroom for secondary school Mathematics education. Results revealed that this approach can help enhance underperforming and high ability students’ Mathematics achievement. Students’ qualitative responses also showed that they benefited from the flipped classroom approach. This is congruent with previous research conducted in higher education settings (e.g., Herreid & Schiller, 2013). Yet teachers should design their flipped classroom according to their students’ ability.

There are several limitations that affect the generalization of our findings and one should exercise caution in interpreting the results of our study. First, due to logistical issues, we could only employ a one-group pre-and-post-test design. Using a more robust design (e.g., randomized experimental design with separate control and intervention groups) could show the effect of flipped classroom on student achievements more clearly. We therefore urge future research to use experimental or quasi-experimental design to examine the effects of flipped classroom. Second, the study sample consisted of Form 6 (Grade 12) students. Future research should examine this approach involving students of other grade levels. Third, the duration of this study ranged from two to four weeks. Conducting a longitudinal study (e.g., one year) can help determine if students’ perceptions of flipped classroom would change over time. Nevertheless, despite these limitations, we believe the findings would benefit other researchers and educators in exploring the use of the First Principles of Instruction to design a flipped classroom teaching and learning approach.

References

Bergmann, J., & Sams, A. (2009). Remixing chemistry Class: Two Colorado teachers make Vodcasts of their lectures to free up class time for hands-on activities. Learning and Leading with Technology, 36(4), 24-27.


Enfield, J. (2013). Looking at the impact of the flipped classroom model of instruction on undergraduate multimedia students at CSUN. TechTrends: Linking Research & Practice to Improve Learning, 57(6), 14-27.


An Action Research Study from Implementing the Flipped Classroom Model in Primary School History Teaching and Learning

Vasiliki Aidinopoulou12 and Demetrios G. Sampson34*

1Greek-French School Jeanne d’Arc, Greece // 2Department of Digital Systems, University of Piraeus, Greece // 3School of Education, Curtin University, Australia // 4Information Technologies Institute, CERTH, Greece // aidinopoulou@gmail.com // demetrios.sampson@curtin.edu.au

*Corresponding author

ABSTRACT

The benefits of the flipped classroom (FC) model in students’ learning are claimed in many recent studies. These benefits are typically accounted to the pedagogically efficient use of classroom time for engaging students in active learning. Although there are several relevant studies for the deployment of the FC model in Science, Technology, Engineering and Maths (STEM) subjects, and at Higher Education and/or High School, there are very few works studying FC in social studies and at primary school level. This paper presents an action research focused on the implementation of the FC model in teaching social studies in primary school. The main scope of this action research, conducted over an entire school year with two different History classes (one representing the experimental group that followed the FC model and the other representing the control group following the traditional lecture based approach) was to compare the use of classroom time for student-centered learning activities and the resulted learning outcomes related to both traditional learning goals of a history course (that is, memorization of historical content) and more ambitious ones such as the cultivation of historical thinking skills (HTS). The study revealed that indeed, the classroom based sessions of the experimental group were used for engaging student-centered activities and that this resulted into better learning outcomes in terms of demonstrating critical HTS. Thus, this initial action research provides encouraging evidences for the potential benefits of the FC model in primary school social studies courses.

Keywords

Flipped classroom model, Primary school, History teaching, Historical thinking skills

Introduction

Traditional teaching of social studies at school level is challenged by both the use of technology and the adoption of inquiry-based teaching strategies in other subjects, such as Science, Technology, Engineering, and Math (STEM) (Bishop & Verleger, 2013; Hwang et al., 2015; Keengwe & Onchwari, 2015). Typically, in traditional history/social studies school curricula, emphasis is given in memorizing large amounts of historical content namely, names, dates and facts, making these courses less attractive to students (Fielding, 2005). Furthermore, still, many history/social studies school teachers adopt traditional teaching strategies, using most of the classroom time for lecturing and assessing students’ ability to memorize content. As a result, students do not actively engage in learning and assessment activities that promote their historical/critical thinking development (Gaughan, 2014). It is also argued that students have common misconceptions about historical knowledge, primary sources, human motivation and historical change which are not easily overcome with traditional teaching strategies (Epstein, 2012). As a result, there are systematic efforts to enhance teaching, learning and assessment of history and social studies at K-12, by exploiting innovative pedagogical designs supported by digital technologies (Lyons, 2008).

On the other hand, the FC model has gained prominence over the past years, as a technology-supported pedagogical innovation which uses classroom time for students to actively engage in interactive learning activities, including personalized feedback and scaffolding from the teacher, while teachers’ traditional lecturing is delivered out of the classroom time with asynchronous video lectures (Chen et al., 2014). Despite the wide take-up of the FC model in Higher Education and STEM subjects (Bishop & Verleger, 2013; Keengwe & Onchwari, 2015; Sergis et al., 2017), there are limited efforts in studying its application in primary school and in history/social studies courses, in particular.

In this paper, it is argued that adopting the FC model in a primary school history course has the potential to use classroom time in a more efficient way, leading to enhanced students’ learning experiences and outcomes. Therefore, the paper reports on the design and results of an action research implemented to investigate this hypothesis and provide evidence on the added value of the FC model.

The remainder of the paper is structured as follows: the Background section presents an overview of the FC model and its implementation in teaching history/social studies school courses. The Action Research Method
section presents the methodology and research questions of the action research. The Educational Design section describes the design of the educational intervention used in the action research. The Results section presents the findings obtained in relation to the two research questions and, finally, the Discussion and Future Work section discusses the lessons learnt and outlines potential future research.

Background

The FC model is an emerging blended learning model widely used both in school and university formal educational settings. Based on sound pedagogical theoretical principles, FC targets to exploit classroom time and space for appropriately designed interactive learning activities differentiated according to individual and group students’ needs (DeLozier & Rhodes, 2016). This section presents an overview of the use of digital technologies in school history teaching and learning, and the implementation of the FC model in K-12 history/social studies.

Technology supported history teaching and learning in K-12

As in all disciplines, the integration of educational technologies in social studies is constantly increasing (Lee & Friedman, 2009). Over the past decade, many history teachers are reported to be hesitant to adopt digital technologies into their classrooms and still rely on the textbook and their own lectures for teaching history (Scheuerell, 2015). Surveys showed that some teachers were skeptical about the effectiveness of technology, while others expressed concerns about the amount of time they had to devote in order to design new activities and material and also the technological competence they should acquire (Townsend, 2010).

On the other hand, research studies have demonstrated the potential of digital technologies in social studies at school level in relation to the development of historical critical thinking skills (Green et al., 2013). To this end, a wide range of digital technologies have been used in history classrooms. Digital libraries, for example, allow teachers and students to freely view and download primary and secondary sources for use in the classroom, providing students with the possibility to examine different perspectives of historical context (Scheuerell, 2015). Simulations, forums, wikis, blogs, documentary film making, social media, whiteboards, web quests and voting technologies are tools that history teachers have been using in an effort to transform the history classroom into a student-centered environment (Haydn et al., 2014). Thus, it is established that K-12 history teaching and learning is now widely exploiting digital technologies.

Using the flipped classroom model in history teaching and learning in K-12

The FC model has been only sporadically implemented in history and social studies at middle and/or high school courses. Typically, in such implementations, students watch video lectures before coming to class. The purpose of these videos is to present the historical content, in addition to the material presented in the course textbook. These video lectures are highly connected with classroom time and are meant to trigger student’s interest for the classroom activities that will follow (Gaughan, 2014). The video lectures are usually made by the teachers with narration, text and enriched with maps, annotations and images. For the creation and distribution of the videos, widely available tools are used, such as iTunes and YouTube (Kotlik, 2014).

On the other hand, the classroom-based face-to-face sessions are used for discussions on the previously watched video lectures, and other student-engaging tasks, such as primary and secondary sources analysis, debating, peer reviewing or simulations (Bergmann & Sams, 2012). From case studies, it is reported that more history teachers tend to adopt the popular flipped-mastery system, according to which, students have to prove mastery in content using their own pace and means of proving academic performance (Bergmann & Sams, 2014).

In history courses that use the FC model, emphasis is given in the cultivation of historical thinking skills. As students interact with material, they learn how to use historical context and perspectives different to their own, towards developing critical historical thinking skills, they analyze and hypothesize, and in general act as historians (Gaughan, 2014; Bergmann & Sams, 2015). Consequently, the memorization of historical content is only a minor learning goal as most goals concern skill development (Fulton, 2014).

Regarding the evaluation of the FC teaching approach in history courses, studies report improvements in students’ learning outcomes both in historical content memorization as well as in the cultivation of historical critical thinking skills. Furthermore, students and teachers demonstrate a positive attitude towards the FC model...
as it increases interest and engagement and makes students more responsible in their studies (Fielding, 2005; Gaughan, 2014; Bergmann & Sams, 2015).

In this context, the present study reports on the design and implementation of an action research in order to provide additional evidence of the impact of the FC model on the exploitation of face-to-face classroom-based teaching time as well as the students’ learning outcomes. The following section outlines the action research method adopted in this work.

**The action research method**

The action research was designed and conducted using McKernan’s model (McKernan, 1991). This model, which is a form of scientific inquiry (McKernan, 1988), meets Lewin’s four original action research phases: plan, act, observe and reflect (Lewin, 1948), but also adds a spiral flow that allows the researcher to reflect on and redesign his action research. It gives the opportunity to the action researcher to go through all the phases again, via a new action cycle and bring additional elements into the study. This section presents the general purpose and research questions of the study as well as the study’s context and characteristics of participants. In addition, the phases of the two action research cycles are described. The section is concluded with the description of the study’s instruments and the data collection and analysis methods.

**Research questions**

The general purpose of the action research was to investigate the extent to which the implementation of the FC model can improve students’ learning outcomes and also lead to better use of classroom time. More specifically, our scope was to compare the use of classroom time for student-centered learning activities and the resulted learning outcomes related to both traditional learning goals of a history course (that is, memorization of historical content) and more ambitious ones such as the cultivation of historical thinking skills (HTS), between two different History classes (one following the FC model and the other following the traditional lecture-based approach). As a result, the following research questions were defined:

- **RQ1.** Does the implementation of the FC model in a primary school history course lead to the use of classroom time for more student-centered activities?
- **RQ2.** Does this contribute to better students’ learning outcomes compared to traditional teaching strategies?

The first research question concerns the exploitation of classroom based teaching time. We argue that in history teaching better use of classroom time is achieved when it accommodates student-centered activities that promote cognitive goals of the higher levels of Bloom’s revised taxonomy such as Analyze, Evaluate and Create (Krathwohl, 2002), as opposed to activities that promote cognitive goals of the lower levels of Bloom’s taxonomy, such as Remembering. The FC model has been attributed with the capacity to allow the teacher to facilitate such learning and assessment activities since it releases time by moving lecture-based content delivery out of the classroom time (Bergman & Sams, 2012). For the second research question to be investigated, the selection of learning goals is required. The specific learning goals of the action research are explicitly described in the “Educational Design” section.

**Study context**

The study was conducted in a two-term-long (24 weeks) history course at primary school. More specifically, the “Roman and Byzantine History” course in the 5th Grade of the Greek primary school national curriculum was the context of this study. The course’s syllabus, in brief, covers a long period of sixteen centuries (146 BC - 1453 AD) from the subjugation of Greece by the Romans to the conquest of Constantinople by the Turks. This includes extensive names, dates and facts that have to be memorized by the students and, moreover, the textbook involves challenging vocabulary in relation to the so far cognitive experiences of students. Thus, a teacher who needs to follow such a syllabus faces all the difficulties outlined in the previous sections. The action research aimed to investigate whether the implementation of the FC model could provide a solution to this educational problem.
Participants

The participants of the study were 49 eleven–year old students who attended the “Roman and Byzantine History” course at the Greek – French “Jeanne d’Arc” Primary School in Greece in two different classes of Grade 5. From these, 26 students (11 boys, 15 girls) formed the experimental group and 23 students (7 boys, 16 girls) formed the control group. These two groups were selected because (a) all students had access to a computer and an internet connection and (b) they demonstrated similar learning achievements in the history courses that they had attended during the previous two school years. Thus, overall, we consider that the selected groups did not present any significant differences in number of students in each group, gender distribution, or prior knowledge that could bias the results. All students were familiar with the use of computers but were unfamiliar with the Moodle (https://moodle.org/) learning management system (LMS) that was used to facilitate the implementation of the FC model. Therefore, classroom time had to be devoted for students to get acquainted with the LMS and their new role in the new teaching strategy.

Procedure

The action research was conducted in two action cycles following the McKernan spiral model as Table 1 depicts. The first cycle took place in four phases. First phase was the Plan. It was conducted from July - August 2014 before the beginning of the first term of the 2014 school year. During this phase, the educational design took place. Second and third phases in the research were Action and Observation. They were completed during the first term from September 11, 2014 until December 10, 2014 when formally the first term ended. The term’s ending was followed by the Christmas holiday season during which the Reflection phase took place. From the Reflection phase, emerged the revised design of the educational intervention. The second cycle of the action research began with the formation of the revised educational design. The revised plan was implemented (Action and Observation) during the second term, from January 8, 2015 until March 10, 2015, when the second term formally ended. Then, followed the reflection of the second cycle and during the period March - June 2015 the data of both terms were gathered and results presented.

<table>
<thead>
<tr>
<th>1st term - 1st action research cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plan</strong></td>
</tr>
<tr>
<td>Selection of resources, creation of new material, design of activity flow and assessment methods.</td>
</tr>
<tr>
<td><strong>Act and Observe</strong></td>
</tr>
<tr>
<td>Implementation of the FC model. Data collection from teacher logs and first term’s assessments.</td>
</tr>
<tr>
<td><strong>Reflect and Re-plan</strong></td>
</tr>
<tr>
<td>Reflection on the first term’s observations and revision of educational design.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd term - 2nd action research cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Act and Observe</strong></td>
</tr>
<tr>
<td>Implementation of the revised plan. Collection of data via teacher logs and second term’s assessments.</td>
</tr>
<tr>
<td><strong>Reflect</strong></td>
</tr>
<tr>
<td>Analysis of both terms’ research data. Presentation of action research results.</td>
</tr>
</tbody>
</table>

Instruments

In order to collect data concerning the use of classroom time, teacher logs were created that documented for every teaching hour, the flow of learning activities and the time spent in each activity. The learning goal of the memorization of the historical content was assessed with the standardized tests of the National curriculum. In order to assess historical thinking skills cultivation, assessment activities were created by the teacher that included open and closed ended questions on primary and secondary sources. For the evaluation of the students’ answers the “Assessing Historical Thinking and Understanding (ARCH): Historical Thinking Skills Rubric” was used (UMBC Center for History Education, 2013). Each assessment question was linked to one learning goal and was evaluated using the rubric’s criteria. The tests were common for both groups and included the same grading scale (from 0 to 10).
Data analysis

Following data collection, the data analysis was divided into the following tasks: (a) analyze the distribution of classroom time in learning activities for the experimental and the control group during the first and second term; (b) analyze the students’ learning outcomes in the historical content memorization assessments for the experimental and the control group in the first and second term; (c) calculate the Mann-Whitney U test and effect size in order to statistically examine the significance of differences in student historical content memorization achievement scores between the experimental and the control group; (d) analyze the student’s learning outcomes in the HTS cultivation assessments for the experimental and the control group in the first and second term; (e) calculate the Mann-Whitney U test and the effect size in order to statistically examine the significance of differences in student HTS cultivation achievement scores between the experimental and the control group; (f) an overall evaluation to identify the benefits and possible challenges of the FC model. Mann-Whitney U tests were employed to investigate for potential significant differences in the assessment scores between the two groups, since the data did not follow a normal distribution and the sample size was not large. The calculation of the effect size was based on the standardized mean difference and evaluated based on Cohen’s d coefficient (Cohen, 1994). Since the effect size provides a quantitative means to measure the “strength” of an observed case, the adopted “strength” intervals are defined as: small (effect size ≤ 0.3), medium (0.3 < effect size ≤ 0.5), large (0.5 < effect size ≤ 0.7) and very large (0.7 < effect size) (Cohen, 1994). These intervals will be exploited during the analysis of the results regarding students’ learning outcomes (RQ2), so as to provide a measure of the impact of the FC model.

Educational design

The purpose of the educational design was to create the educational intervention to be used in the action research. This section describes the five phases (analysis, design, development, implementation, evaluation) of the ADDIE model (Dick, Carey & Carey, 2001) as applied to the study.

Analysis Phase: During this phase, the educational problem that primary school history courses face was identified. Matters such as the emphasis in historical content memorization as the prime learning outcome, insufficient teaching time for engaging students in learning activities that support historical thinking skills cultivation, lack of student interest and common student historical misconceptions, raised the need for a new approach that could improve history teaching and learning. The analysis phase also included the identification of the experimental and control groups’ characteristics as described in the previous section, as well as the students’ learning environment characteristics.

Design Phase: During this phase, learning goals were defined and the flow of activities and assessment methods were selected. The learning goals of the intervention were formed after studying the following international standards for teaching history: (a) the National Standards for Social Studies Teachers (Myers, 2000), (b) the Common Core State Standards for History/Social Studies - Grade 6-8 and (c) the UCLA Historical Thinking Standards (Common Core State Standards Initiative, 2010). These learning goals were common for both research groups and they were grouped into categories and subcategories:

- **Category A: Historical content memorization**, namely, to remember and recall names, dates and facts from the long-term memory.
- **Category B: Historical Thinking skills cultivation**:  
  - **Understanding the concept of time**, namely, to distinguish between past, present and future time, to identify and use the temporal structure in a historical narrative, to measure and calculate calendar time, to interpret data presented in time lines and create time lines.
  - **Understanding historical sources**, namely, to identify the author of the historical document, to identify the historical facts of the source, to differentiate between historical facts and historical interpretations, to draw upon data in historical maps, visual, literary and musical sources.
  - **Historical analysis and interpretation**, namely, to analyze cause-and-effect relationships, to argue using historical evidence, to hypothesize on how different historical actions could have let to different results, to formulate historical questions, to evaluate the actions and decisions of historical persons based on their results.

Apart from the learning goals, the flow of activities is elaborated in the implementation phase. The selected assessment methods have been described in the previous section.
Development Phase: During this phase, new material was created to successfully support the implementation of the FC approach. First of all, a series of video lectures was created that presented the content material. The creation of videos followed standard design principles (Koumi, 2006) with the typical duration being 9-10 minutes and the lectures usually being divided into three parts (3 minutes each). The video lectures and all new content were provided to students via the Moodle LMS. In addition, activities were developed by exploiting Moodle, such as quizzes, wikis and forums for students to study and interact. All the above contributed to the application of the FC model.

Implementation Phase: During this phase, the educational intervention was implemented. The procedure of the intervention is presented in Table 2 and then further elaborated.

<table>
<thead>
<tr>
<th>Table 2. Educational intervention procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
</tr>
<tr>
<td><strong>Before Class</strong></td>
</tr>
<tr>
<td>Studied material using the Moodle LMS</td>
</tr>
<tr>
<td>Interacted at the forum</td>
</tr>
<tr>
<td><strong>During Class</strong></td>
</tr>
<tr>
<td>Lectures on the new historical content</td>
</tr>
<tr>
<td>Q&amp;A session with the teacher</td>
</tr>
<tr>
<td>Teacher’s feedback</td>
</tr>
<tr>
<td>HTS cultivation activities (if time was available)</td>
</tr>
<tr>
<td><strong>After Class</strong></td>
</tr>
<tr>
<td>Studied material using the textbook and all digital resources except from the video lectures</td>
</tr>
<tr>
<td>Interacted at the forum</td>
</tr>
</tbody>
</table>

The students of the experimental group prior to attending classroom activities studied the video lectures in their online digital classroom that was implemented using the Moodle LMS. They interacted at the forum by publishing questions, comments or answering to their classmates’ questions. They also answered quizzes based on what they had previously studied. For those who had the time and interest, extra material was provided, relevant to the video lectures. During the face-to-face classroom sessions, first, the students were involved in Q&A activities with the teacher in order to solve any questions that remained unsolved. The teacher then assessed their historical content memorization by verbally asking questions. The assessment was followed by the teacher’s feedback on students’ answers. The remaining teaching time was spent in engaging with historical thinking skills (HTS) cultivation activities, such as collaborative activities and debates. The teacher observed the whole process and gave directions and feedback. At the end of each unit the assessments of the two main categories of learning goals took place. Students completed standardized tests with open and closed ended questions referring to historical content memorization and HTS cultivation. The assessments were followed by the teacher’s feedback on students’ learning outcomes and the students were rewarded with digital badges.

The students of the control group during the face-to-face classroom sessions were exposed to traditional lectures on the new historical content. After the lectures, Q&A sessions with the teacher followed in order to respond to questions that were created during the delivery of the new content. The remaining classroom time was invested in engaging with HTS cultivation activities. In most cases though, the remaining time was not sufficient. When the students of the control group returned home, they had to study and practice using their textbook and the historical content that was delivered to them. They could also use digital resources (articles, images, presentations) that were published in the LMS (the control group had the same access to resources as the experimental group except for the video lectures). If they had any further questions they could publish them in the forum. Following their home-based individual study, they attended classroom-based follow-up sessions which started with Q&A activities and then the teacher assessed their historical content memorization and HTS cultivation (same as in the experimental group). The assessment was followed by the teacher’s feedback on students’ answers and the delivery of the new historical content. At the end of each unit the assessments of the two main categories of learning goals took place. Students completed standardized tests which were common to the experimental group.

Evaluation Phase: During this phase, the completed educational intervention was evaluated. All data collected from the implementation phase were processed and the results were presented.
Results

Results regarding the use of classroom time (RQ1)

The purpose of the study was to examine the benefits of the FC model relating to the use of classroom time and the students’ learning outcomes. This section summarizes the results that corresponded to the two research questions. Table 3 presents the distribution of classroom time in learning activities for the experimental and the control group in the first and second term.

<table>
<thead>
<tr>
<th>Table 3. Distribution of classroom time in learning activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>1st term</td>
</tr>
<tr>
<td>Control group</td>
</tr>
<tr>
<td>Experimental group</td>
</tr>
<tr>
<td>2nd term</td>
</tr>
<tr>
<td>Control group</td>
</tr>
<tr>
<td>Experimental group</td>
</tr>
</tbody>
</table>

As Table 3 depicts, lecture was the main activity that the control group spent most classroom time in, whereas the experimental group spent no time in this activity. This is due to the implementation of the FC approach in the experimental group that transferred lectures entirely to the individual learning space. The control group also spent more time in verbal assessment of historical content memorization than the experimental group. The increase of classroom time spent by the experimental group in verbal assessment of historical content memorization in the second term is due to the revision of educational design, as a result of teacher’s reflection to first term teaching and learning analysis. It was observed that the students of the experimental group who were less often verbally assessed on historical content memorization were not as well prepared as the students of the control group that were more often verbally assessed on historical content memorization. For that reason, in the second action research cycle, the teacher spent more time assessing that learning goal in the experimental group.

Moreover, in the first term, the control group spent more classroom time in Q&A sessions and feedback than the experimental group. This is due to the fact that the experimental group students participated more actively in the Moodle forum, asking most of their questions before coming to class. The students of the control group preferred to ask questions and, hence, received feedback at the time when the new content was delivered to them, during classroom sessions. In the second term though, the classroom time spent in Q&A sessions and feedback by the experimental group was increased, because the time spent in assessment of historical content memorization had also increased. Therefore, new questions came up especially from students that came poorly prepared to the class.

A significant difference of classroom time invested by the experimental group in HTS cultivation activities is observed in the first term. This is due to the implementation of the FC model that released classroom time and allowed students to engage in HTS cultivation activities. This finding provides evidence to positively answer our first Research Question, namely that the FC model allows classroom time to be more effectively used, because it is dedicated in student-centered engaging activities that facilitate the development of higher cognitive goals such as HTS cultivation. The decrease in classroom time spent by the experimental group in HTS cultivation activities in the second term is due to (a) the revision of the educational design as described above and (b) the fact that teaching hours became fewer in the experimental group compared to the first term because of school schedule (field trips etc.). Finally, the two study groups spent equal classroom time in standardized tests, which were common for both groups.

Results regarding to students’ learning outcomes (RQ2)

As shown in Figure 1 and Table 4, during the first term, the experimental group had an increased improvement in students’ score means. More specifically, the group reported (compared each time to the previous assessment) (a) in the 1st formative assessment a 4.38% increase and (b) in the 2nd formative assessment, a 1.7% increase. Correspondingly, the control group had an increase followed by a decrease in students’ score means. The group
reported (a) in the 1st formative assessment a 3% increase, and (b) in the 2nd formative assessment, a 1.5% decrease.

![Graph showing comparison between Control and Experimental groups]

In the second term, the experimental group and the control group reported similar performances. The experimental group reported in the 4th formative assessment a decrease of 8.4%, compared to the previous assessment. This decline may be due to increased difficulty of the certain standardized test. In the summative assessment there was a 2.98% increase. The control group also had in the 4th formative assessment a 3.8% decrease and a 0.5% increase in the summative assessment.

Table 4. Historical content memorization: Mann-Whitney U test and Effect Size results

<table>
<thead>
<tr>
<th>Assessment of historical content memorization</th>
<th>Control group [N = 23]</th>
<th>Experimental group [N = 26]</th>
<th>U</th>
<th>p</th>
<th>Effect Size (ES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Assessment</td>
<td>8.54 (1.49)</td>
<td>8.45 (1.41)</td>
<td>277.5</td>
<td>.672</td>
<td>0.06</td>
</tr>
<tr>
<td>1st Formative Assessment</td>
<td>8.80 (1.39)</td>
<td>8.82 (1.36)</td>
<td>293.5</td>
<td>.912</td>
<td>0.02</td>
</tr>
<tr>
<td>2nd Formative Assessment</td>
<td>8.67 (1.54)</td>
<td>8.97 (0.93)</td>
<td>295.0</td>
<td>.936</td>
<td>0.25</td>
</tr>
<tr>
<td>3rd Formative Assessment</td>
<td>9.01 (0.95)</td>
<td>9.15 (0.93)</td>
<td>272.0</td>
<td>.587</td>
<td>0.15</td>
</tr>
<tr>
<td>4th Formative Assessment</td>
<td>8.67 (1.48)</td>
<td>8.38 (1.34)</td>
<td>251.0</td>
<td>.335</td>
<td>0.21</td>
</tr>
<tr>
<td>Summative Assessment</td>
<td>8.63 (1.59)</td>
<td>8.63 (1.53)</td>
<td>280.0</td>
<td>.702</td>
<td>0.003</td>
</tr>
</tbody>
</table>

As the Table 4 depicts, no statistical significance was evident between groups. However, it is evident from the data that the experimental group had in general the same or higher achievement scores in historical content memorization than the control group. There is only one different observation, the fourth formative assessment, which is considered an isolated event. The calculation of the effect size further attests to this conclusion, since it depicts a low effect in the differences in student historical content memorization achievement scores between the groups. Therefore, the data suggest that the FC model had a limited positive impact on enhancing students’ content memorization capacity.

Table 5 depicts the results of the Mann-Whitney U test based on the assessment scores for each student group in terms of the three Historical Thinking skills categories. As the Table 5 depicts, regarding “Understanding the concept of time,” no statistically significant difference exists in the diagnostic test, therefore the two groups had similar prior skill. However, the data for the remaining three tests signify that the experimental group showed a statistically significant improvement compared to the control group. The effect size supports this finding, attributing medium and large effects to the exploitation of the FC model. In a similar vein, the data for the “Understanding historical sources” category attest to (a) the initial similar skill level of the two groups (no statistical significance exists), and (b) to the statistical significant improvement of the experimental group in the subsequent three assessment tests. The large effect size results again corroborate these findings. Finally, in the “Historical analysis and interpretation” category, a similar pattern of results is observed, namely the two groups showed uniform skill levels during the diagnostic test, but subsequent statistically significant differences in favor of the experimental group (with corresponding medium to large effect sizes).

Overall, the analysis of data signifies that the implementation of the FC model had a significant positive impact on developing students’ HTS, but not so regarding their historical content memorization capacity. This is an
interesting finding which provides evidence on the added value of the FC model, namely its potential for facilitating the development of students’ historical thinking skills.

Table 5. HTS cultivation: Mann-Whitney U test and effect size results

<table>
<thead>
<tr>
<th>Assessment of historical thinking skills</th>
<th>Mean (SD)</th>
<th>U</th>
<th>p</th>
<th>Effect Size (ES)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control group [N = 23]</td>
<td>Experimental group [N = 26]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding the concept of time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic Assessment</td>
<td>5.33 (1.74)</td>
<td>5.58 (1.91)</td>
<td>273.5</td>
<td>.580</td>
</tr>
<tr>
<td>1st Formative Assessment</td>
<td>5.87 (2.08)</td>
<td>7.02 (1.73)</td>
<td>217.5</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>2nd Formative Assessment</td>
<td>5.98 (1.54)</td>
<td>6.92 (0.93)</td>
<td>215.0</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Summative Assessment</td>
<td>6.20 (1.59)</td>
<td>7.12 (1.53)</td>
<td>225.0</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Understanding historical sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic Assessment</td>
<td>5.22 (1.83)</td>
<td>5.58 (1.91)</td>
<td>266.0</td>
<td>.477</td>
</tr>
<tr>
<td>1st Formative Assessment</td>
<td>5.54 (1.89)</td>
<td>6.73 (1.54)</td>
<td>203.0</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>2nd Formative Assessment</td>
<td>5.65 (1.72)</td>
<td>6.54 (1.24)</td>
<td>217.0</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Summative Assessment</td>
<td>5.76 (1.76)</td>
<td>6.83 (1.51)</td>
<td>210.0</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Historical analysis and interpretation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic Assessment</td>
<td>5.00 (1.69)</td>
<td>5.38 (1.83)</td>
<td>263.0</td>
<td>.430</td>
</tr>
<tr>
<td>1st Formative Assessment</td>
<td>5.98 (2.10)</td>
<td>6.92 (1.29)</td>
<td>208.0</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>2nd Formative Assessment</td>
<td>6.09 (1.29)</td>
<td>7.02 (1.53)</td>
<td>204.5</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Summative Assessment</td>
<td>6.30 (2.24)</td>
<td>7.12 (1.53)</td>
<td>205.5</td>
<td>&lt; .05</td>
</tr>
</tbody>
</table>

Discussion and future work

Benefits of the flipped classroom model

Overall, we can conclude that the FC model contributed to a more effective use of classroom time since it released time from lecturing to more engaging student-centered learning activities. Consequently, in the experimental group, more classroom time was available for implementing learning activities that promote learning goals relevant to HTS.

Regarding students’ learning outcomes, the data showed that the FC model had a low positive impact on developing students’ traditional historical content memorization achievement scores between the two groups. This means that both study groups had similar learning outcomes in historical content memorization although the students of the control group spent significantly more classroom time with teacher’s lectures compared to the students of the experimental group who spent no time for live lectures during classroom-based face-to-face sessions. This observation can provide promising research hypotheses for future investigations:

- The teacher’s lectures in the control group may not have been engaging enough in order to increase students’ learning outcomes in historical content memorization.
- The video lectures developed for the experimental group may have been as engaging as needed in order to promote learning outcomes in historical content memorization similar to the teacher’s live lectures.

Regarding students’ HTS, however, there was a clear positive impact of the FC model. More specifically, the experimental group showcased better learning outcomes than the students of the control group (in a statistically significant manner and with medium to large effect sizes). If we take under consideration that the HTS cultivation is a process that takes place mainly inside the classroom, it is reasonable to assume that there is a cause and effect relationship between the classroom time spent on HTS cultivation and the learning outcomes of HTS cultivation. In other words, the reason that the students of the experimental group had better learning outcomes than the students of the control group, may strongly be the fact that the FC model facilitated more classroom time to be invested on HTS cultivation activities.

Overall, potential future work could include additional longitudinal action research studies to provide additional evidence on the impact of the FC model on students’ performance and learning experiences. Furthermore, these works could also adopt a more granulated level of analysis for studying and revealing how the FC impacts the learning patterns (e.g., learning material study patterns or engagement patterns) of individual students.
Challenges of the flipped classroom model

Aside from the benefits of the FC model that are described above, there are certain challenges in its use that need to be taken into consideration. First of all, it is most likely that students are not acquainted with the FC model and the different learning roles and activities it introduces. This can be even more relevant in school environments where the traditional teaching model is predominantly used. As a result, before implementing the FC model, the teacher should devote teaching time in order to "train" the students. Training students on the way they should study history or any other subject matter may prevent future obstacles in the learning process.

Furthermore, there is a need for parents to be familiarized with the FC model especially in primary school where the family has an important role in student’s learning and can contribute to or obstruct the model’s efficiency.

The process of creating a course based on the FC model requires significant time invested on behalf of the teacher and also some level of technological competence. That is, the teacher needs to develop learning activities and resources for the classroom. In addition, it was observed that some students of the experimental group were not self-motivated enough and they did not thoroughly prepare themselves by watching and interacting with the video lectures nor did they study the material in general. The result was that they came to class unprepared. This led to less participation and engagement from those students who consequently fell behind. Therefore, it is important that the teacher is empowered with tools that facilitate him in monitoring individual students’ home-based preparation activities and collect data that can help him re-design activities based on differentiated instruction principles. Overall, if the FC model is to be widely adopted, tools for Teaching and Learning Analytics (Sergis & Sampson, In Press) are required to support teachers as reflective practitioners.

Acknowledgements

The work presented in this paper was conducted as part of the first author’s MSc Diploma Thesis at the MSc Programme “Digital Systems and Services (Track on e-Learning)” of the Department of Digital Systems, University of Piraeus, Greece and it has been partly supported by the State Scholarship Foundation (IKY) under the grant No 30275/2013B (IKY Fellowships of excellence for postgraduate studies in Greece – Siemens Programme).

References


Gaughan, J. E. (2014). The Flipped classroom in world history. *The History Teacher, 47*(2), 221-244.


Conceptualizing “Homework” in Flipped Mathematics Classes

Zandra de Araujo1*, Samuel Otten1 and Salih Birisci1,2

1University of Missouri, Columbia, MO, USA // 2Artvin Coruh University, Seyitler, Turkey // dearaaujoz@mizzou.edu // ottensa@mizzou.edu // birisci@gmail.com

*Corresponding author

ABSTRACT

Flipped instruction is becoming more common in the United States, particularly in mathematics classes. One of the defining characteristics of this increasingly popular instructional format is the homework teachers assign. In contrast to traditional mathematics classes in which homework consists of problem sets, homework in flipped classes often takes the form of an instructional video. This paper presents a framework for flipped mathematics homework that categorizes types of homework and also draws on technology literature and mathematics education literature to discern quality for each type. For each category of homework, we provide illustrative examples from a study of flipped mathematics implementations. We show how the quality of instructional videos can vary and that, in addition to instructional videos, teachers also assigned videos to motivate subsequent in-class work on mathematics tasks. We close with directions for future work.

Keywords

Flipped instruction, Mathematics education, Teaching, Homework, Multimedia

Introduction

Within flipped instruction, teachers invert or “flip” the settings in which teacher lecture and student practice occur. In the last few years, teachers’ implementation of flipped instruction has increased dramatically. Flipped instruction has now reached the point where more than two-thirds of teachers in the United States report flipping a lesson, if not the entire course (Smith, 2014). Although flipped instruction originated in the context of university-level sciences (Mazur, 1991), it has become especially popular in the context of secondary and post-secondary mathematics (Moore, Gillett, & Steele, 2014; Zack et al., 2015). The increase in flipped mathematics courses is perhaps due to the recent proliferation of online mathematics instructional videos (e.g., Khan Academy) and the longstanding tradition, particularly in the United States, of mathematics instruction characterized by lecture and exposition (Stigler & Hiebert, 1999). This lecture-based approach translates well to video-based delivery. Further, the more frequent use of lectures at the secondary and post-secondary levels than at the elementary level (Banilower et al., 2013) corresponds with greater utilization of flipped instruction in those settings (Yarbro, Arfstrom, McKnight, & McKnight, 2014). Thus, secondary and post-secondary mathematics classes in the United States constitutes a ripe site for empirical investigations of flipped instruction.

The potential innovative power of flipped instruction can stem from videos themselves, from a different use of class time, or from a synergy between the two. Because research on flipped instruction is still limited (Uzunboylu & Karagozlu, 2015), we focus on the most defining component of flipped instruction: the instructional video. Teachers in flipped classrooms typically assign the instructional videos for homework. Thus, the videos replace the problem sets that are the typical form of homework in mathematics. We present a framework for flipped mathematics homework that categorizes types. We also draw on technology literature and mathematics education literature to discern quality for each type.

Literature review

Much of the existing literature related to flipped instruction involves expositional descriptions of how authors intend to implement it (e.g., Bergmann & Sams, 2012) but does not involve empirical, third-party observations of those implementations. In mathematics education, research is emerging but the early studies tend to focus on small numbers of flipped classes and on outcomes rather than on the nuances of the implementations. For example, Clark (2015) examined two flipped algebra classes and compared their performance on unit content tests and their survey responses to those of students in non-flipped algebra classes. Clark found no significant differences with respect to performance and only moderate differences in the survey results, with slight preference indicated toward flipped instruction. This study, however, did not give attention to the specific ways in which flipped instruction was implemented. Further, the study did not account for the fact that other implementations of flipped instruction could vary widely from Clark’s. Similarly, DeSantis and colleagues (2015) compared one instance of flipped geometry instruction with non-flipped instruction and found no significant differences with regard to mathematical performance. They did find a slight difference in student
opinions, in this case with students preferring non-flipped instruction. Others (Love, Hodge, Grandgenett, & Swift, 2014; Sahin, Cavlazoglu, & Zeytuncu, 2015) found more positive results for flipped instruction in university mathematics courses. However, as with the previously cited studies, the details and variability of flipped instruction were not accounted for in the study design. This may be because there is not yet a robust framework for flipped instruction in mathematics.

To work toward such a framework, we looked beyond studies of flipped instruction itself and considered work related to individual aspects of flipped instruction. One important aspect is the way class time is used during flipped implementations. However, in this article we focus on another important aspect: the flipped homework that sets the stage for in-class activity. Because flipped homework is typically video or multimedia content, one might develop a framework based on the holistic level of whether the video contains a recorded lecture (with minimal post-production), a voice-over presentation (with audio synchronized to slides or on-screen writing), or a picture-in-picture presentation (with extensive post-production). Ilioudi and colleagues (2013), for example, found that, over three weeks, mathematics learning correlated more positively with recorded lecture videos than with voice-over lectures.

Although this format-focused approach is worthwhile, we sought a framework that incorporated the full instructional triangle—teacher, students, and content (Cohen, Raudenbush, & Ball, 2003)—with respect to video or multimedia homework. The instructional triangle highlights that teaching is not just what teachers do, but what they do with students in relation to content. The typical environment for this is in a classroom, but the triangle also provides a way to think about instruction in a home environment with respect to flipped videos or multimedia. The videos contain content and the teacher is present, either virtually by being a voice in the video or indirectly by choosing the video to assign. The students are also present, either watching the video or possibly interacting with it.

To consider each of these facets in more detail, we viewed the content through the lens of the Mathematical Quality of Instruction (MQI) instrument (Learning Mathematics for Teaching Project, 2011). The MQI instrument provides a well-established way to consider the coherence and richness of the expressed ideas. The MQI utilizes a coding framework based on presence/absence and appropriateness/inappropriateness of several features of mathematics instruction. For instructional videos, we found that the MQI’s “richness and development of the mathematics,” “language,” and “unmitigated mathematical errors” sub-categories were of particular relevance because the other sub-categories (e.g., “responding to students”) do not relate directly to instructional videos or multimedia assigned as homework. Because the MQI was applied to video-recorded lessons, the application to flipped lecture videos seems an appropriate use of the instrument.

For the teacher, we chose to consider some of the decisions the teacher made (consciously or unconsciously) in crafting the video. We used Clark and Mayer’s (2008) six principles of design for instructional materials with embedded multimedia content. The multimedia principle states that video developers should wisely and judiciously select and add graphics to text. The contiguity principle states that relevant text should appear near those graphics. The modality principle states that graphics should be explained with audio. Relatedly, the redundancy principle states that audio that simply reads written text aloud should not be included. The coherence principle states that graphics and audio should be used only if pertinent. Finally, the personalization principle advises that developers use a conversational tone when possible. Taken together, these principles can be used to assess design aspects of videos or multimedia assigned as flipped homework. They are analogous to many of the choices a teacher might make in a classroom setting (e.g., what to write on the board and when, tone and formality of speech).

Finally, we considered the students with respect to their opportunities for interaction. What is the extent to which the flipped homework incorporates interactive digital elements, such as embedded questions or other features that require students to interact rather than only watch or listen? Within mathematics education specifically, Moyer-Packenham and Westenskow (2013) defined virtual manipulatives as interactive, dynamic representations that may support students in developing mathematical understanding. For example, teachers might provide students with applets featuring interactive graphs that allow students to see the impact changes to coefficients have on the graph of a particular function. Moyer-Packenham and Westenskow’s (2013) review of over 60 research reports found virtual manipulatives to have potentially large positive effects on students’ mathematical learning, when used in conjunction with well-designed curricula. They concluded that the use of virtual manipulatives provides students with opportunities to engage in mathematics that differs from those presented by paper-and-pencil activities. These findings indicate virtual manipulatives may be worth including in a framework for flipped mathematics homework. Other opportunities for interaction might be more general, such as questions or hints embedded in the videos.
The literature above relates to videos as a means for presenting mathematical content. A flipped mathematics class, however, may also incorporate videos that are not lectures. Instead, videos may be used to motivate ideas that will be subsequently explored in class. Jackson and colleagues (2013) highlighted in general the importance of setting up a mathematical task for students. In the case of flipped mathematics classrooms, the set up might occur at least in part through the homework video. In a later section, when we present our coding scheme for homework videos, we further elaborate the ways in which we conceptualized distinctions among homework that is assigned to set up subsequent mathematics activity.

Overall, there is a large amount of existing scholarship that relates to particular aspects of flipped homework but there are not yet accepted frameworks that coordinate these aspects in coherent ways. The purpose of this article is to describe a framework of flipped mathematics homework and illustrate it with data from a project investigating the implementation of flipped mathematics instruction at the secondary and post-secondary levels.

Method

We use the term homework to describe the work that is assigned for students to complete at home. In flipped instruction, a lecture or expository video or multimedia presentation is typically assigned to be viewed as homework. In our analysis, we considered all manner of video or multimedia presentations (e.g., combinations of animations, audio, or text) the teachers assigned to students. However, in this study the assignments most commonly consisted of video recordings. Therefore, we will refer to these assignments as “video” for the remainder of the paper.

We examined four U.S. teachers’ classrooms—Dr. Moore’s college-level calculus class, Ms. Schaefer’s college-level college algebra class, Mr. Forrest’s high-school level pre-calculus class, and Ms. Temple’s eighth grade mathematics class. Dr. Moore, Ms. Schaefer, and Ms. Temple had chosen to flip nearly all of their lessons, whereas Mr. Forrest continued to use a traditional instructional model but included flipped lessons occasionally. For each teacher we observed several lessons, collected copies of all materials (both in class and at home) associated with those lessons, and conducted pre- and post-lesson teacher interviews. We also collected survey data from the teachers about their implementation and perspective on flipped instruction.

![Homework in a Flipped Mathematics Class](image)

For this particular study, we conducted a two-phase analytic process, constituting our framework for flipped mathematics homework. For the first phase, we categorized the homework according to the teacher’s purpose for selecting the video—lecture or set-up/motivation. We considered these two categories separately because the goal for each differed. The goal for a lecture video (including worked examples) is to deliver content or demonstrate a process. In contrast, a set-up/motivation video is not created for the purpose of content delivery but to establish a situation or context for subsequent mathematical activity in class. Because we wanted a framework that captured the purpose and quality of the homework, we decided to separate these two categories.
This is particularly important because the field of mathematics education has established differing criteria for effective teaching practice for each of these purposes.

We considered not only the videos but also teacher interview and survey data in determining the purpose of the assigned videos. For a particular lesson, it is also possible that the teacher assigns no homework at all. It is also possible that a teacher assigns a set of problems or exercises to be completed individually by the students—that is, a typical non-flipped homework assignment. We recognize that teachers using flipped instruction may not flip every lesson. Thus, the inclusion of these additional categories allows us to capture differences between a teacher who flips 100% of their lessons and a teacher who flips many lessons but still incorporates some non-flipped homework assignments or perhaps no homework assignments. Figure 1 contains all of these possibilities for homework in general. However, we focus the remainder of this paper on the video homework because this is characteristic of flipped lessons specifically.

For the second phase of analysis, we assessed the quality of each video in its respective category. This entailed three members of the research team individually coding each video. The full team, five members, then met to resolve any disagreements that resulted from the individual coding. This aspect of our framework is important because it recognizes that the way in which one flips (e.g., the quality of the videos) may be more important than simply whether one flips. Our operationalization of the notion of quality is described in the following sections, wherein we provide illustrative examples of the various types of homework in our framework.

Lecture videos

Each video was coded as either a lecture video or a set-up/motivation video. In mathematics, these lecture videos typically include explanations of mathematical terms or ideas, justifications of mathematical propositions, or worked examples of mathematics problems. They may be created by the teacher or drawn from an existing source such as Khan Academy or educational YouTube channels. For example, a video of a teacher explaining the idea of the y-intercept of linear functions would fall into our category of lecture videos, as would an applet with accompanying text that demonstrates how to solve a certain type of algebra problem. This type of video/multimedia is prevalent in secondary mathematics. This may be due to these videos’ alignment with traditional notions of mathematics instruction as involving a teacher explaining ideas to students and showing them what they need to be able to do (Smith, 1996; Stigler & Hiebert, 1999).

Determining the quality

To assess the quality of lecture videos we developed a coding scheme based on the literature reviewed above, particularly portions of the Mathematics Quality of Instruction (MQI) instrument (Learning Mathematics for Teaching Project, 2011), Clark and Mayer (2008), and Moyer-Packenham and Westenskow (2013). Thus, our framework includes three key criteria for high quality lecture video in mathematics:

- it is coded as present-appropriate in the “richness and development of the mathematics” and “language” sub-categories of the MQI and coded as “not present” in the “unmitigated mathematical errors” sub-category of the MQI;
- it adheres to at least five of the six principles of digital material design (Clark & Mayer, 2008); and
- it incorporates virtual manipulatives (Moyer-Packenham & Westenskow, 2013), dynamic representations of mathematical concepts relevant to the goals of the lesson, or digital interactive features (e.g., quizzes, applets, discussion boards).

Note that these criteria are conceptually separate, not integrated, but it is worthwhile to consider them together because they all contribute to the instructional triangle (Cohen, Raudenbush, & Ball, 2003).

A lecture video can satisfy zero, one, two, or all three criteria. For example, a video could violate most of the digital design principles, not incorporate a virtual manipulative, but still have richness and development of mathematics and appropriate language as defined in the MQI. Although coding was based primarily on the video/multimedia itself, future video analyses may require examination of subsequent lesson observation to inform the coding. This subsequent analysis could uncover things such as whether a mathematical error in a video is addressed explicitly by the teacher in class. In the following sections we provide examples from mathematics classes of lecture videos that satisfy different criteria.
Example of video satisfying one criterion

This first example video is from a flipped, college algebra class taught at a community college by Ms. Schaefer. For homework, Ms. Schaefer typically created her own video lectures using Explain Everything (Version 3.21, available from http://itunes.apple.com), a screen casting application for iPads. Explain Everything allows the user to use the iPad as a whiteboard while simultaneously capturing the user’s speech in a single video. Ms. Schaefer stated that she did not script the videos prior to recording them. She also noted that the videos took a lot of time to develop. Ms. Schaefer described her expectation for students as they watched the videos.

Well and I told them it, it’s basically as though I were talking to you. If you were in class, what would you be doing? Taking notes, writing things down, important things that I say, same concept. It is just instead of me talking now, I’m talking at home at you. So that’s, that’s really what we’re looking to do there.

We selected a lecture video Ms. Schaefer created to explain how to graph rational functions. This video is illustrative of a typical lecture video in Ms. Schaefer’s classroom (Figure 2).

In this 12-and-a-half-minute video, Ms. Schaefer began the video by defining rational functions as those that are written as a fraction in which the variable \( x \) is in the denominator. Then, Ms. Schaefer explained how to determine the domain of a rational function and then defined asymptotes as imaginary lines that graphs approach but may not cross. For the remainder of the video, Ms. Schaefer described each of the three types of asymptotes (vertical, horizontal and slant (oblique)) and provided examples of rational functions that included each type. During the solution process she used algebraic representations without accompanying graphs or other representations. Although the video focused on graphing rational functions, the sole graph included in the video was a depiction of a vertical asymptote absent the accompanying curve for the rational function. At several points while she worked through examples of how to determine asymptotes she encouraged students to pause the video and check their answers. She took advantage of the multi-colored pen feature of the Explain Everything application to highlight important points.

We assessed Ms. Schaefer’s video as meeting one of the quality criteria for lecture videos—Clark and Mayer’s (2008) design principles for digital materials. Ms. Schaefer kept a conversational tone throughout much of the video (personalization principle) and the ways in which she incorporated text and audio satisfied the multimedia, contiguity, coherence, and modality principles. She did not meet the redundancy principle because there were a number of instances in which she read aloud text on the screen. Although the video did not have mathematical
errors and made use of proper mathematical terminology, the richness and development of the mathematical idea was not highly rated. Therefore, the video did not meet the third criteria related the mathematical content of the lecture. Further, Ms. Schaefer did not include interactive features or virtual manipulatives in her video and so did not meet the second criteria for interactive features.

**Example of video satisfying two criteria**

Dr. Moore taught a flipped calculus I course at a university. Like Ms. Schaefer, Dr. Moore typically developed lecture videos for his students and often consisted solely of an instructional video. Dr. Moore’s videos captured his written work with a video camera rather than screencasts (Figure 3). The video we examined focused on the First Derivative Test and provides a typical example of the videos Dr. Moore created.

![Figure 3. Excerpt from Dr. Moore’s video](image)

We analyzed a 22-minute video focused on the First Derivative Test. The video began by identifying and illustrating key terms such as relative extrema. Throughout the video, Dr. Moore referenced events from in-class sessions and used different colors and drawings to highlight important points. Dr. Moore included a number of examples to illustrate the concepts he discussed. For example, when discussing tangent lines with slopes of zero, he drew two such examples on the graph (Figure 3). After discussing the purpose for using the First Derivative Test, Dr. Moore gave several worked examples of how to use the test.

We determined that Dr. Moore’s video met two of the quality criteria. Like Ms. Schaefer, Dr. Moore’s video adhered to the multimedia, contiguity, modality, and coherence principles. Although the video did not generally use a conversational tone (personalization principle), Dr. Moore avoided much of the redundancy of reading written text aloud. Thus, the five principles that were addressed were sufficient to meet the first criteria for high quality videos. Dr. Moore’s video was free of mathematical errors and the mathematical language was appropriate. Further, in contrast to Ms. Schaefer’s video, Dr. Moore included a rich mathematical explanation of why the First Derivative Test works. Therefore, the video met the third criteria. Like Ms. Schaefer’s video, Dr. Moore’s video did not include any virtual manipulatives or interactive features and therefore did not meet the second criteria.

**Example of video satisfying three criteria**

Our final example comes from Ms. Temple’s flipped eighth grade mathematics class. Ms. Temple created her lecture videos in an iBook format. This format allowed her to not only provide students with videos but also
include text, examples, practice problems, and embedded quizzes. The homework we considered was from a chapter on scatterplots. In assessing Ms. Temple’s homework, we considered all the features of the iBook shown in Figure 4. In addition to the iBook, Ms. Temple provided her students with a set of guided notes to complete. The guided notes included references to particular pages of the textbook, had questions for the students to answer, and indicated when and how students were to engage with the iBook. Thus, students used the notes to guide them through the entirety of the homework which drew on both the iBook and textbook. Despite the variety of resources Ms. Temple provided students for homework, we categorized this assignment as a lecture video because her purpose was to instruct students on the topic of scatterplots.

**Figure 4.** A screenshot from Ms. Temple’s homework assignment

The first section of the iBook chapter included a set of five guiding questions for the students to consider as they worked through the homework. Ms. Temple also included examples of scatterplots with positive, negative, and no correlation. The students were then to watch a video. Ms. Temple created a six-and-a-half-minute video to further explain what a scatterplot is and how to create one. The video featured screencasts of Ms. Temple’s written work on an iPad with the accompanying audio. The video began by providing students with the correct solutions to the questions on the guided notes and encouraged students to pause the video to check their answers (Figure 5). Then, Ms. Temple provided a worked example of how to create a scatterplot and a line of best fit. Throughout the example, she emphasized conventions for labeling the scatterplot’s axes and identified common student errors. She then provided a problem for students to complete and asked students to pause the video to complete it. She closed the video by providing the solution for the practice problem. Following the video, the students were to complete a set of review questions on the iBook to check their understanding. The final component of the homework was an online quiz for the students to complete, the results of which Ms. Temple could retrieve online.

We assessed Ms. Temple’s homework as meeting all three of the criteria for high quality lecture videos. Ms. Temple’s video, like the videos from Ms. Schaefer and Dr. Moore, met the first criteria for design principles of digital media. Unlike the others, however, Ms. Temple’s video included all six principles. For example, she avoided reading text that was provided in graphic form (redundancy principle) and described given graphics with audio (modality principle). Like Dr. Moore, Ms. Temple used proper mathematical language and did not have mathematical errors in her video. Although video did not include a rich development of the mathematical idea,
the textbook section students were instructed to read did. Thus, the video fully addressed the third criteria. Finally, in contrast to the other two videos, Ms. Temple’s iBook did include interactive features (the review questions and practice problems), satisfying the second criteria.

Summary

In looking across the three video examples we see three different levels of quality. Though both Ms. Temple and Dr. Moore’s videos included well developed mathematical explanations, proper use of language, and were free of errors, Ms. Temple’s inclusion of interactive features led to the video’s classification as higher in quality than Dr. Moore’s. These interactive features help to engage learners in mathematical activity rather than passively viewing the videos. We posit that, based on extant literature (e.g., Moyer-Packenham & Westenskow, 2013), the inclusion of these interactive aspects may provide greater opportunities for student learning and so videos with these features can be assessed as higher in quality than those absent them.
In contrast to both Dr. Moore and Ms. Temple’s videos, Ms. Schaefer’s video focused solely on procedures and did not motivate why or how the procedure for determining asymptotes works. This in conjunction with the video’s lack of digital manipulatives or interactive features led to this particular video assessed as lower in quality than the other two videos because it met just one of the criteria. This assessment indicates there may be less opportunities for students to learn and engage in mathematics as they view the video. For example, though students may gain some familiarity with how to determine an asymptote for a particular function, the video does not indicate how or why this procedure is needed. Further, students may finish watching the video and have little understanding of asymptotes as features of a particular function rather than an independent graph. The difference in content and user experiences presented in these examples motivate the need to consider criteria for moving beyond top level categories such as lecture videos into criteria to describe the quality of these videos.

Set-up/Motivation videos

Not all videos assigned as homework in flipped instruction are expository in nature. In some cases, a teacher assigns students a video that poses a mathematical problem or establishes a non-mathematical context motivating the work that will occur during the subsequent class period. We call these set-up/motivation videos, and they occurred less frequently in our data than lecture videos.

Determining the quality

There are no existing coding schemes specifically designed to identify the effective features of these types of videos in mathematics. There is, however, research on the process of setting up mathematical tasks in class. For example, Jackson and colleagues (2013) examined how teachers use the set-up to establish common language, introduce the context of the task, or review mathematical relationships that will be pertinent. They argued that these set-up practices have implications for subsequent learning opportunities. González and Eli (2015) also studied the tensions between providing enough set-up for students to get started but not so much that it reduces the cognitive demand of the task.

This existing research depends largely on the actions of the teacher and so the analytic approaches do not apply directly to set-up videos themselves. Thus we devised a new coding scheme for assessing the quality of set-up videos. In this scheme, the video is considered together with any questions or activities that are also assigned as part of the homework:

- it explicitly prompts a mathematical problem or idea that will be explored or developed as part of the lesson’s learning goals;
- it implicitly contains a mathematical problem or idea that will be explored or developed as part of the lesson’s learning goals; or
- it does not contain a discernible mathematical problem or idea that will be explored or developed as part of the lesson’s learning goals.

Such videos, by definition, do not contain a full lecture or exposition of mathematical content. However, these codes describe the extent to which they prompt or prepare the students for the mathematical activity that will occur in class. A video is of the explicit type if one can watch the video (and read any associated homework questions, such as what to attend to in the video) and predict with confidence the mathematical learning goal of the subsequent in-class lesson. Such a video may intellectually motivate a purely mathematical question or it may present a mathematical idea in an applied context. In either case, the mathematical idea of interest is explicit and clear. A video of the implicit type contains mathematical ideas that will serve the lesson’s learning goals. In contrast to explicit types, however, the learning goals are not obvious from simply watching the video itself. They may only be clear in retrospect after seeing the subsequent in-class activity. The final type of set-up video is one that does not contain mathematical ideas relevant to the lesson’s learning goals but perhaps only relates to a real-world context. We further illustrate each of these three types of set-up video in the next sub-sections.

**Example of a video that implicitly connects to a learning goal**

Mr. Forrest was a high school mathematics teacher. His teaching approach involved students being “the drivers of the mathematics” because they “retain more when they ‘discover’ concepts on their own versus having [the teacher] deliver it to them.” Most of his lessons were not flipped but consisted of in-class investigations,
followed by independent practice as homework. He did, however, flip some individual lessons. The video we examined came from his pre-calculus class and focused on the use of trigonometric functions to determine unknown components of triangles. The video shows a plane landing at an airport that is just behind a public beach in Sint Maarten (see Figure 6). Mr. Forrest asked the students to watch the video and write down questions they wanted to have answered about the situation.

![A screenshot from Mr. Forrest’s assigned video](image)

_Figure 6. A screenshot from Mr. Forrest’s assigned video_

In the video, there are many mathematical and non-mathematical phenomena to which the students could attend. Moreover, the directions accompanying the video allowed the students to generate any question rather than focusing them on formulating questions that can be answered with trigonometric functions. Many students did ask questions appropriate to the trigonometric goals of the subsequent in-class work. For example students asked about the height of the plane above the beach, the height of the plane above the chain-link fence, and the distance the man in the hat is from the runway. However, several other students asked questions that did not align with the goals, such as how loud the plane was or how many planes land there each day.

Overall, this video establishes a context for trigonometric work. It implicitly contains mathematical relationships relevant to the lesson goals, such as the angle of descent formed between the runway and the plane’s motion. However, Mr. Forrest, not the video, was the one to guide students in identifying the key mathematical relationships. Further, he had to clarify which of the students’ many questions were appropriate with regard to the lesson goals. The video also did not explicitly contain the required information needed to answer the questions. Mr. Forrest, for example, directed students to use the internet to look up typical airplane landing angles because it could not be measured accurately from the video.

**Other examples**

Because of the rarity of set-up videos in our data, we briefly mention other examples drawn from external sources. Meyer (2011) wrote about creating or selecting videos in which the desired mathematical question is clear or even unavoidable. Figure 7 shows an image that aligns with our notion of explicit prompting because a natural question is whether the basketball will go through the hoop. Furthermore, because the video was shot from a fixed position and the relevant features are in view (e.g., the basketball’s path, the hoop, the ground), the video lends itself for direct use in the subsequent in-class activity related to quadratic functions.

Finally, a hypothetical example of a video that has no discernible connection to the lesson’s mathematical goals would be a video clip from a popular movie assigned as homework because the subsequent lesson will involve analyzing movie box office statistics. In this case, the video is meant to excite students or connect to their natural interest in movies but it does not actually contain any box office statistics or any content that will be central to the lesson’s learning goals. In other words, it focuses purely on the context of subsequent in-class activities, not
the mathematical ideas. Such a video would be analogous to some of the non-mathematical marginal images that are included in U.S. mathematics textbooks as purely visual elements or as opportunities to connect with students’ non-mathematical interests; such marginal images are absent from the textbooks of some other countries, such as Singapore (Erbas, Alacaci, & Bulut, 2012).

Figure 7. An image from an explicit video example

Summary

Set-up/motivational videos can be more or less explicit about the learning goals that follow. Nevertheless, this category of video constitutes a substantial departure from typical non-flipped homework because it is about preparing for mathematical work rather than practicing a skill that has already been taught. These videos, however, constitute only the starting point of the actual set-up to the mathematical work. Mr. Forrest, for instance, still had more set-up work to do in class as he took the students’ questions and formulated them into a specific trigonometric task. Teachers may also, to greater or lesser degrees, engage in the kinds of set-up practices that Jackson and colleagues (2013) identified, such as establishing common language. Hence this category may be particularly challenging to analyze in isolation from how it is used subsequently in class.

Conclusion

We presented a framework that distinguishes between various purposes and quality of homework in flipped mathematics classes. This framework provides a basis for design research focused on developing effective materials for flipped instruction in mathematics. Development of flipped materials is already occurring at an increasing rate but is not yet informed by empirical evidence. By establishing shared conceptions of various types of homework and homework quality, researchers can move forward in conducting investigations of how these various forms impact student learning, if at all.

Because this framework was developed from a relatively small data set, future work is needed to understand whether the framework can adequately categorize teachers’ homework practices in flipped classrooms. Furthermore, additional work is required to understand whether the proposed quality indicators impact student learning. Thus, although we have developed this framework based on literature on mathematics teaching in non-flipped environments and our sample of flipped classrooms, future work is needed to further refine this framework.

In our ongoing work, we have thus far found that teachers seldom included interactive features in their lecture videos. We have also yet to find videos that included the dynamic virtual manipulatives described by Moyer-Packenham and Westenskow (2013). We posit this may be due to the teachers’ lack of familiarity with such tools or the time required to embed interactive features into videos. We also found that, by a wide margin, the teachers we studied more frequently assigned lecture videos than set-up/motivation videos. Further study is needed to understand whether the type of homework has differing impact on student learning and whether the proportions we found were similar to those found in the wider population of teachers.
Additionally, further study is needed to examine the quality criteria we have proposed. In particular, it is not known whether certain criteria are more important than other criteria in terms of the impact homework has on student learning. For example, could the mathematical features of a lecture video have a greater impact on student learning than the digital design principles, or vice versa? Future work might bring members of the educational technology and mathematics education communities together to better understand if integration of the criteria is possible and/or necessary. For example, is there a way to integrate the MQI with features of digital design, rather than consider them separately as we have done? Although these questions remain unanswered, the framework presented provides a first step and common conception from which to devise such studies.

In future investigations of homework in flipped classroom, student behavior is an important mediating factor to consider. Student behavior refers to the extent to which students actually watch the videos, or not, and to students’ behaviors while watching the videos (pausing, skipping ahead, multitasking, etc.). Although the work at home can directly influence what occurs in class, it is possible that student behaviors can greatly mitigate this influence. For example, if a high-quality lecture video is assigned but no students watch it, the video is unlikely to have a strong relationship with the in-class activity or the students’ learning. Finding ways to measure and categorize these behaviors is crucial for those who wish to examine connections between flipped instruction and student learning outcomes.

Looking beyond homework, it is likely that the in-class implementation of flipped instruction is just as important or more important than the homework. For example, after students watch a set-up video as homework, a teacher might organize an engaging and cognitively-demanding task that builds on the video or might simply ask students to complete a number of low-level tasks loosely associated with the video. These different in-class choices may be more important than the quality of the homework. In other words, the real potential of flipped instruction may not lie in the videos but in what the flipped format allows the teacher and students to do together in the classroom.

Understanding flipped instruction is crucial because although a majority of teachers have implemented this model of instruction, little is known about how teachers are implementing it, it is also unclear what impact flipped instruction may have on student learning. The existing literature base is predominantly expositional or anecdotal and is often written in ways that suggest flipped instruction is a unified teaching model. Our work suggests that flipped instruction varies greatly and this variance must be attended to in order to draw meaningful conclusion. Our framework draws distinctions among homework types that could set the foundation for more nuanced investigations of flipped learning. This may in turn help the research base keep better pace with implementation.

Acknowledgements

This work was funded by the University of Missouri’s ReSTEM Institute. We thank Nicole Fyten, Abigail Heffern, Milan Sherman, and the participating teachers and students.

References


Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. Eugene, OR: International Society for Technology in Education.


Investigating the Potential of the Flipped Classroom Model in K-12 ICT Teaching and Learning: An Action Research Study

Christoforos Kostaris¹², Stylianos Sergis²⁴, Demetrios G. Sampson³⁴*, Michail N. Giannakos⁵ and Lina Pelliccione³

¹¹⁵th Junior High School of Peristeri, Athens, Greece // ²Department of Digital Systems, University of Piraeus, Greece // ³School of Education, Curtin University, Australia // ⁴Information Technologies Institute, CERTH, Greece // ⁵Department of Computer and Information Science, Norwegian University of Science and Technology, Norway // xxkostaris@yahoo.gr // steliossergis@iti.gr // dametrios.sampson@curtin.edu.au // michailg@idi.ntnu.no // L.Pelliccione@curtin.edu.au

*Corresponding author

ABSTRACT

The emerging Flipped Classroom approach has been widely used to enhance teaching practices in many subject domains and educational levels, reporting promising results for enhancing student learning experiences. However, despite this encouraging body of research, the subject domain of Information and Communication Technologies (ICT) teaching at K-12 education has not yet been explicitly researched. This is a considerable shortcoming given (a) the globally acknowledged need to effectively design and deliver ICT curricula to students at all levels of school education (primary, junior and senior secondary) towards cultivation of digital competences, and (b) the recurring research evidence that the Flipped Classroom approach can enhance students’ learning experiences in Science, Technology, Engineering and Math (STEM) K-12 education. Therefore, the contribution of this paper is the design and implementation of an action research for studying the effect of the Flipped Classroom approach in K-12 ICT teaching and learning. The action research employed a quasi-experimental design using experimental-control groups, from two classes (a total of 46 students) for an entire semester of the school year following the National (Greek) Curriculum in junior high school ICT studies. The results of this study provide evidence for potential advantages in students’ cognitive learning outcomes related to subject domain knowledge, the exploitation of teaching time during the classroom face-to-face sessions, the students’ level of motivation, as well as their level of engagement.

Keywords

Flipped classroom model, K-12 education, ICT curriculum teaching, Teaching time, Student learning outcome, Student motivation, Student engagement

Introduction

Information and Communication Technologies (ICT) curricula are a core priority of educational policies worldwide, for cultivating students’ diverse digital competences (Commonwealth of Australia, 2010; US Department of Education, 2010; Eurydice, 2012). These digital competences are considered a core strand of the 21st century skillset and industry work-force requirements are shifted towards attaining and exploiting such competences (OECD, 2014). Thus, ICT educational curricula have been increasingly required to effectively cultivate student digital competences by promoting more student-centred teaching approaches (e.g., Balanskat & Engelhardt, 2015). For example, the project-based approach has been positively evaluated in the context of K-12 education as a promising method to effectively engage students in the learning process and to develop their digital competences (e.g., Wang et al., 2016; Tsai et al., 2015).

Furthermore, on top of exploiting such student-centered teaching approaches, educational innovations such as the Flipped Classroom Model (FCM) have also been studied for improving learning experiences and competences. The FCM has been proposed as a method to maximize the use of teacher-supported face-to-face classroom-based sessions towards delivering hands-on activities and individual scaffolding. To achieve this, it opts for substituting the traditional teachers’ lecture with appropriately designed educational resources which can be engaged by the student in an autonomous manner (Bergmann & Sams, 2012).

The FCM has received a growing level of research attention given the promising results that showcase its capacity to enhance teaching practice and deliver (among others) better students’ cognitive learning outcomes and motivation in a wide range of subjects and educational levels, including ICT and STEM (Giannakos et al., 2014). However, regarding ICT teaching, existing FCM research has been solely addressed in Higher Education, where a large body of research exists that provides positive evidence on the potential of FCM to provide enhanced student experiences. Despite this evidence, however, no explicit focus has yet been placed on evaluating the FCM in the context of K-12 ICT teaching. This presents an important challenge, especially considering the range of successful implementations of the FCM in other compulsory school education subjects.
Thus, the contribution of this paper is to address the aforementioned issue and evaluate the FCM within the context of K-12 ICT curriculum. More specifically, the work presents the design and implementation of an action research for evaluating the capacity of the FCM to enhance students’ learning experiences from a range of perspectives, namely (a) cognitive learning outcomes, (b) distribution of different types of learning activities during the face-to-face sessions, (c) levels of motivation during the learning activities, and (d) level of students’ engagement during the learning activities.

The remainder of the paper is structured as follows. The Background section defines the FCM and describes existing research work on its implementation in Higher Education to enhance students’ learning experiences. The Research Methodology section presents the design and methodology of the action research. The Results section presents the findings of the action research in terms of the four research questions defined. Finally, the Discussion and Future Work section presents the conclusions drawn and outlines potential strands of further research.

Background

The Flipped Classroom Model (FCM) is an emerging blended learning model, which argues for improving the student-centered exploitation of face-to-face sessions, by minimizing teacher lecture and increasing students’ active learning and collaboration (Bergmann & Sams, 2012). Teacher-facilitated face-to-face sessions can provide students with unique learning experiences through direct access to both their classmates (for collaborative activities) as well as to feedback and scaffolding by their teacher (DeLozier & Rhodes, 2016). Therefore, optimizing the exploitation of face-to-face sessions is very important towards creating more student-engaging learning experiences. To accomplish this, the FCM argues that lectures can be substituted with appropriately designed and/or selected learning materials (e.g., educational videos) which can be provided to the students for prior “home study” before the face-to-face session, in a more autonomous manner (Chen et al., 2014). Thus, classroom-based time can be directed at student-engaging learning activities.

The FCM has received a growing level of research attention in a wide range of subjects and educational levels, such as STEM and ICT (Bishop & Verleger, 2013). This growing interest is fueled by the benefits attributed to the FCM primarily in terms of enhancing cognitive learning outcomes (e.g., Love et al., 2014; Kong, 2014) and motivation (Baeppler et al., 2014; Hung, 2015). Surprisingly, however, despite the significant body of research for evaluating FCM in other K-12 subjects, ICT as a subject domain has received no explicit research attention. This is a significant shortcoming since K-12 ICT curricula are now revisited and reformed to meet the demands for diverse digital competences for all students.

On the other hand, FCM has been investigated in the context of Computing Curricula in Higher Education, so as to enhance students’ cognitive learning outcomes. More specifically, Gannod et al. (2008) and Amresh et al. (2013) presented positive results on using FCM in software engineering courses. In the same vein, Mason et al. (2013) investigated the FCM in an advanced ICT engineering course with a specific focus on problem solving. Hayashi et al. (2014) reported that the FCM had a positive impact on students’ outcomes on programming skills. Finally, Horton et al. (2014), Reza and Baig (2015) and Jonsson (2015) also reported improvements in the cognitive learning outcomes of students in FCM-supported experimental groups compared to their control counterparts.

Apart from cognitive learning outcomes, research in FCM in Higher Education ICT teaching has also focused on students’ level of motivation and/or engagement. More specifically, Day and Foley (2006) reported positive effects of FCM on enhancing students’ motivation during an ICT course (also on their cognitive learning outcomes). Similarly, Herold et al. (2012) utilized interviews of students and instructors and highlighted that the FCM enhanced students’ motivation and quantity of discussions during the sessions. Dolgopolovas et al. (2014) reported on a significant increase in students’ motivation in a FCM-enhanced computer science course, which was attributed to the higher proportion of hands-on activities delivered in the face-to-face sessions (even though no explicit modelling of these activities was performed). These findings are consistent with Sureka et al. (2013), and Gehring & Peddycord (2013), who also reported enhanced student motivation in FCM-enhanced computer science courses. Davies et al. (2013) positively evaluated the impact of the FCM to enhance student motivation, satisfaction as well as cognitive learning outcomes in an introductory course on spreadsheets (linking these findings to the time spent on different types of learning activities). Finally, Choi (2013) studied the FCM in an introductory software engineering course and concluded that the experimental group showed improvement in collaboration skills and cognitive learning outcomes.
Overall, existing works highlight a consistent pattern of findings that argue in favor of the capacity of the FCM to enhance Higher Education ICT teaching in terms of students’ cognitive learning outcomes, motivation, engagement and more student-oriented exploitation of face-to-face sessions. These findings are further complemented by the students’ positive perceptions on the added value of FCM to both cultivate their skills related to ICT curriculum (e.g., Tanner & Scott, 2015) as well as to enhance the overall learning experience (e.g., Fryling et al., 2015).

Therefore, based on (a) the rich body of evidence on the capacity of the FCM to enhance students’ learning experiences in diverse K-12 subjects (Rahman et al., 2014; Giannakos et al., 2014), and (b) the current lack of any such evidence in relation to K-12 ICT teaching despite the promising relevant findings in the Higher Education context, a research challenge is identified. This challenge relates to investigating the potential of the FCM model to provide similar improvements in the learning experiences of students in the novel context of K-12 ICT teaching, especially considering the importance of this subject for cultivating necessary digital competences for students. Furthermore, a more holistic conceptualization of “learning experiences” is adopted by exploiting all four learning experience indicators identified in existing works (but not used in combination in any of them). This more holistic approach is important to study the impact of the FCM from a perspective that involves both student-related dimensions (namely, cognitive learning outcomes, level of student motivation, level of student engagement) and teaching-related dimensions (the types of learning activities in face-to-face classroom-based sessions).

Therefore, towards addressing this research challenge, the paper describes the design and implementation of an action research in order to evaluate the capacity of the FCM to enhance students’ learning experiences in a K-12 ICT course. More specifically, the action research focused on evaluating the added value of the FCM in terms of (a) improving students’ cognitive learning outcomes, (b) re-distributing the types of learning activities held during the face-to-face sessions towards including more student-centered activities, (c) enhancing students’ motivation, and (d) enhancing students’ engagement.

Research methodology

Action research

Action research is defined as a form of data-driven disciplined inquiry in which a practitioner (namely teacher-researcher) aims to understand, analyze and, potentially improve their practice (Cohen et al., 2007). In the context of this work, the aim was related to investigating the capacity of the FCM model to enhance ICT teaching practice in K-12 curriculum.

The action research of this work was designed following the widely used four-phase process of Lewin (1948), namely Plan, Act, Observe and Reflect. More specifically, the Plan phase related to the design of the action research in terms of research questions, methodology and evaluation protocol. Additionally, it included the educational design of the ICT course which would be delivered to the experimental and control groups. The Act and Observe phases were correspondingly addressed at implementing the action research following the methodology adopted and collecting students’ educational data. Finally, the Reflect phase was related to analyzing the educational data collected towards answering the defined research questions.

In order to ensure study trustworthiness, a set of four relevant criteria were adopted from Shenton (2004):

- **Credibility (internal validity)** and **Confirmability (objectivity)** criteria were addressed by: (a) adopting well-established research methods and tools for data collection and data analysis, (b) triangulation of findings derived from diverse data sources, (c) promoting peer-scrutiny of the research process, and (d) building on previous research findings (defined in the “Background” section).

- **Transferability (external validity)** criterion was addressed by: (a) clearly defining the design aspects of the study (e.g., sample size and characteristics, data collection and analysis methods /tools and study time period), and (b) identifying and reporting study limitations.

- **Dependability (reliability)** criterion was addressed by: (a) defining and reporting on the research design and implementation methodology, (b) defining and reporting on data collection and analysis processes, and (c) appraising the study, by reporting on the results to answer specific Research Questions and elicit insights for practice and research.

Each of these criteria is further elaborated in the following sections.
Research questions

Based on the insights from the literature analysis reported in the “Background” section, the following research questions were defined:

- **RQ1:** Does the exploitation of the FCM in an ICT K-12 course (in particular, junior high school) lead to enhanced students’ cognitive learning outcomes related to the subject domain compared to a traditional teaching control group?

- **RQ2:** Does the exploitation of the FCM in an ICT K-12 course (in particular, junior high school) affect the distributions of learning activity types delivered in the face-to-face sessions compared to a traditional teaching control group?

- **RQ3:** Does the exploitation of the FCM in an ICT K-12 course (in particular, junior high school) lead to enhanced students’ motivation compared to a traditional teaching control group?

- **RQ4:** Does the exploitation of the FCM in an ICT K-12 course (in particular, junior high school) lead to enhanced students’ engagement in learning activities compared to a traditional teaching control group?

Participants and context of study

The study was conducted over a period of a full semester (8 weeks), in the context of the 2nd grade of junior high school ICT studies (equivalent to 8th grade, 14 years old) within the Greek National Curriculum. The course covered introductory concepts of computing including computer hardware components and their interrelation, basic software design principles as well as information processing.

The action research was implemented with the full population of two classes (i.e., 46 students), one as the experimental group (which attended the FCM-enhanced ICT course) and the other as the control group (which attended the “traditional” ICT course). Each of the student groups comprised 23 students, with 11 boys and 12 girls.

Approval and consent for conducting the action research was requested and granted by (a) the school leader (with communication to the Ministry of Education), and (b) the parents of the students selected. The students were informed that they were allowed to quit their participation at any time (also removing the relevant data collected). Finally, anonymization of all student data collected was performed throughout the implementation of the action research.

Procedure

The action research cycle comprised four phases. The first phase (Plan) was conducted from June - August 2013 prior to the start of the school year. The focus of this phase (apart from delineating the action research methodology) was focused on the educational design of the ICT course (for both the experimental and the control group). Regarding the experimental group, the Plan phase also included (a) the development and/or selection of additional learning material mainly in the form of educational videos, which would be provided to the students for their home-based study prior to the face-to-face session, and (b) the design and configuration of the online classroom environment, which would be used for delivering the learning activities outside the physical classroom. The online classroom was implemented using the Moodle Learning Management System (https://moodle.org/).

Both instantiations of the ICT course (for the experimental and the control group) were designed following the highest possible similarity in the educational design elements, in order to minimize the possibility of biased results due to vastly different teaching approaches. More specifically, the main teaching approach adopted in both groups was a project-based approach, in which students worked collaboratively to define, implement and present tangible deliverables related to the scope of each unit of the course. Furthermore, in both groups the following teaching techniques were exploited (a) the Jigsaw technique, for fostering active collaboration of students (Tahir et al., 2011) and (b) Web-quests, for promoting students’ “hands-on” practice and active investigation. The assessment methods for both groups included written assessment tests (which contributed to their final grade) and collaborative project deliverables. Furthermore, the educational objectives and face-to-face session frequency and duration were similar for both student groups. The core difference between the two student groups was related to the incorporation of the FCM standpoints in terms of learning activity distribution between the face-to-face classroom-based sessions and the “homework” sessions, as follows.
Regarding the control group, the flow of learning activities for each week was initiated with the face-to-face session, in which the practitioner presented the new learning material/concepts (mainly through lecture), prior to any other learning activity. During the remaining time, the students engaged in their (collaborative) project-based activities. After each face-to-face session, the students were assigned homework to be completed at home. Regarding the experimental group, the flow of learning activities for each week included a set of learning activities, prior to the face-to-face session. During this Moodle-based pre-session, the students were primarily engaged in studying the educational material (e.g., educational videos) provided by the practitioner and self-assessment quizzes. This process was addressed at familiarizing students with the basic concepts of the follow-up face-to-face session. Therefore, the upcoming face-to-face session could be directed on (collaborative) project-based activities and feedback provision.

Following the Plan phase, the Act and Observe phases were triggered, coinciding with the 8 school weeks in which the ICT course was delivered to both student groups. Finally, the Reflect phase was implemented, capitalizing on the outcomes and data collected from the two previous phases, lasting two months.

**Instruments**

For the scope of this action research, data from diverse sources were collected in order to allow for effective triangulation of findings (Phillips & Carr, 2010). For all instruments used, content- and construct- validity was ensured by the practitioner as well as “external reviewers”, namely (a) the “critical friend” of the practitioner (see below), and (b) the researchers (other authors) supporting the practitioner during the action research, who are experts in the field of computer science, educational research and educational technologies. Furthermore, these “external reviewers” also provided critical feedback on (a) interpreting the data collected by the practitioner (e.g., journal observations or survey data), and (b) eliciting insights from data analysis for answering the Research Questions.

Regarding RQ1, the students’ cognitive learning outcomes were assessed through four assessment tests, common for both student groups, delivered at regular time intervals. An initial diagnostic test was delivered prior to the beginning of the action research aiming to assess students’ prior knowledge from the previous grade. The test contained closed questions (true - false, multiple choice), open-ended questions (short answer) as well as the development of a concept map. The results of this initial diagnostic test were used to divide each student group in three performance-based clusters, namely low-medium-high performers (Table 1).

<table>
<thead>
<tr>
<th>Categories of performance</th>
<th>Control group</th>
<th>Group size (N)</th>
<th>Experimental group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Performers</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Medium Performers</td>
<td>8</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>High Performers</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

This clustering (further discussed in the “Results” section) was exploited during the data analysis, to offer more granulated insights on the impact of the FCM in each performance-based cluster. The remaining three tests were correspondingly delivered after the end of the second week, after the end of the sixth week and, finally, after the eighth week of the course. All tests were assessed in a 20-point scale, following the Greek National Curriculum standards.

Regarding RQ2, the teacher journal technique (Altrichter et al., 2008) was employed, to capture the types of activities that the students engaged with during the face-to-face sessions. The types of learning activities followed a custom typology comprising five activity types, namely teacher lecture, student-teacher interactions, student-student collaboration, “hands-on” competence-building activity and assessment activities (i.e., both standardized tests as well formative assessment). Each face-to-face session was analysed in terms of the time spent on each activity type (which were not mutually exclusive). Furthermore, the teacher journal was also populated with the practitioners’ observations on students’ performance as well as the observations of a “critical friend”, who was a senior peer practitioner.

Regarding RQ3, students’ motivation was measured using the standardized Instructional Materials Motivation Survey (IMMS) questionnaire (Keller, 2010). The IMMS questionnaire incorporates 5-point scale questions based on the ARCS motivation model, which classifies students’ motivation in four dimensions, namely Attention, Relevance, Confidence and Satisfaction. More specifically, Attention is related to intriguing students’
interest during the teaching process. Relevance refers to the perceived level of meaningfulness that the teaching process and material has for the students, whereas Confidence refers to the expectations of the students for succeeding in the course. Finally, Satisfaction is related to the positive attitudes of the student towards the teaching practice. The IMMS questionnaire was selected since it is a widely used and validated instrument (e.g., di Serio et al., 2013). For this study, the internal consistency reliability of the IMMS questionnaire was positively evaluated using the Cronbach’s alpha coefficient (0.88 < α < 0.93 for the four dimensions; α = 0.96 for the overall instrument).

Finally, regarding RQ4, a custom engagement rubric was formulated. This rubric comprised nine 5-point Likert items, aiming to assess students’ level of engagement in the face-to-face sessions. The engagement rubric was populated by the practitioner, based on their teaching journals at the end of each week. The internal consistency reliability of the instrument was positively evaluated using the Cronbach’s alpha coefficient (α = 0.88).

Data analysis

Regarding RQ1, the students’ assessment scores on each of the three standardized tests were compared across the two groups using paired-sample and independent sample t-tests and ANCOVA tests. Furthermore, the assessment scores in each group were analysed against the within-group performance-based clusters in order to elicit how the FCM impacts students of different performance levels (i.e., low-medium-high performers).

Regarding RQ2, for each student group the exploitation of the face-to-face teaching time was analysed in terms of the typology of learning activity types using descriptive statistics. The analysis was aimed to elicit to what extent the experimental group engaged in different types of learning activities.

Regarding RQ3, the two student groups were compared using independent sample t-tests, to identify statistically significant differences in their level of motivation (based on the IMMS instrument).

Finally, regarding RQ4, independent sample t-tests were employed, to elicit statistically significant differences among the two groups in terms of engagement. Furthermore, data analysis for the within-group performance-based clusters was performed, to investigate the effect of the FCM on each cluster.

All the aforementioned data analysis tasks were performed using the IBM “Statistical Package for the Social Sciences” (SPSS), version 22 for Windows.

Results

Results regarding students’ cognitive learning outcomes (RQ1)

The RQ1 aimed to investigate the added-value of the FCM on the students’ cognitive learning outcomes. Figure 1 presents the results from the analysis of the students’ assessment tests (mean values) for each student group.

![Figure 1. Results of students’ standardized assessment tests](image-url)
As the Figure 1 depicts, the student groups were similar in terms of prior cognitive knowledge. Beyond that point, however, the experimental group clearly outperforms the control group in all three remaining assessment tests. In order to examine whether there is a statistical significant difference, independent-sample *t*-tests were implemented for each standardized assessment test. The results are depicted in Table 2. Levene’s test for equality of variances was calculated in each case and showed equal group variances.

<table>
<thead>
<tr>
<th>Standardized assessment tests</th>
<th>Control group [N = 23]</th>
<th>Experimental group [N = 23]</th>
<th><em>t</em>(df)</th>
<th><em>p</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Test</td>
<td>15 (1.8)</td>
<td>15.3 (1.75)</td>
<td>-7.49 (44)</td>
<td>.458</td>
</tr>
<tr>
<td>Assessment Test #1</td>
<td>15.7 (2.20)</td>
<td>16.8 (1.51)</td>
<td>-1.877 (44)</td>
<td>.067</td>
</tr>
<tr>
<td>Assessment Test #2</td>
<td>15.7 (1.99)</td>
<td>18.3 (1.25)</td>
<td>-5.320 (44)</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Assessment Test #3</td>
<td>16.9 (1.32)</td>
<td>18.1 (1.25)</td>
<td>-3.315 (44)</td>
<td>&lt; .05</td>
</tr>
</tbody>
</table>

As the Table 2 depicts, there is no statistically significant difference between the groups’ assessment scores on the diagnostic test, therefore the two groups were similar in this regard. Furthermore, regarding the assessment test #1, the difference between the control and experimental groups has increased from 0.3 to 1.1 points, although this difference is not considered statistically significant (Table 2). Regarding the two final assessment tests, the results of the experimental group are statistically significant compared to the control group. Thus, as the course was progressing, the benefits of the FCM were increasing on the students’ performance. In order to eliminate the potential of prior knowledge bias in the results, ANCOVAs were additionally performed to test whether the cognitive learning outcomes between groups were improved in a statistically significant manner, accounting for the students’ grades in the diagnostic test. The assumption of homogeneity of regression was verified in all cases. The ANCOVA results again verified the previous findings and showed statistically significance between groups for all three tests, namely \([F(1,43) = 3.832, p < .05]\) for assessment test #1, \([F(1,43) = 55.973, p < .00]\) for assessment test #2 and \([F(1,43) = 16.432, p < .00]\) for assessment test #3.

In order to further enhance the robustness of the results, the improvement of both groups was studied in terms of their corresponding diagnostic and final test results using paired-samples *t*-tests. The results showed that both student groups had a statistically significant improvement in their cognitive learning outcomes, namely for the experimental group \((t(df) = -9.668(22); p < .00)\) and the control group \((t(df) = -9.665(22), p < .00)\). Therefore, the instance of the ICT course delivered to the control group was also efficient in improving the students’ cognitive learning outcomes. Thus, the statistically significant difference in the improvement of the experimental group over the control group can be attributed to the exploitation of the FCM.

Finally, the data were analysed in terms of the within-group, performance-based student clusters of the experimental group. The findings are depicted in Figure 2, in terms of the percentage of improvement in the assessment scores between the initial diagnostic test and test #3.

As the Figure 2 depicts, the FCM was most beneficial for the low-performing students. This is an important finding that provides evidence for the added value of the FCM on ICT teaching to foster the improvement of students that are experiencing performance-related difficulties.
Results regarding the distribution of learning activity types in face-to-face sessions (RQ2)

The RQ2 aimed to study how the teaching time of face-to-face sessions was exploited for each group and evaluate whether the FCM can actually facilitate teachers to promote more student-centred practices. Figure 3 presents the results obtained, which are depicted as the overall mean value for the all 8 weeks.

![Figure 3. Percentage of learning/assessment activity frequency in face-to-face sessions for each student group](image)

As Figure 3 depicts, teachers’ lecture was the primary learning activity of the control group. This is expected since all new learning content was delivered by the teacher in-class. The control group also invested a considerable amount of teaching time in student-teacher interactions as well as on “hands-on” activities. This is due to the student-centered teaching techniques which were employed for both student groups. Student-student collaboration, however, received a very low frequency percentage, which can be considered a significant shortcoming. Finally, assessment activities for the control group were mainly related to the standardized assessment tests. On the contrary, the largest fraction of teaching time for the experimental group was invested in student-teacher interactions and student-student collaboration. This supports the standpoint that FCM can enable teachers to better exploit their teaching time. Finally, “hands-on” activity implementation was also very frequently implemented during face-to-face sessions, as well as formative assessment/feedback provision tasks (which complemented the standardized tests).

Results regarding students’ level of motivation (RQ3)

The RQ3 aimed to investigate whether the exploitation of the FCM would lead to enhanced students’ motivation. Table 3 presents the results obtained in terms of the four dimensions of the IMMS questionnaire.

<table>
<thead>
<tr>
<th>Motivation dimension</th>
<th>Mean (SD) Control group [N = 23]</th>
<th>Mean (SD) Experimental group [N = 23]</th>
<th>t(df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention</td>
<td>3.6 (0.32)</td>
<td>4.5 (0.24)</td>
<td>10.526 (44)</td>
<td>&lt; .00</td>
</tr>
<tr>
<td>Relevance</td>
<td>3.5 (0.28)</td>
<td>4.25 (0.32)</td>
<td>9.183 (44)</td>
<td>&lt; .00</td>
</tr>
<tr>
<td>Confidence</td>
<td>3.69 (0.35)</td>
<td>4.40 (0.24)</td>
<td>7.984 (44)</td>
<td>&lt; .00</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>3.62 (0.42)</td>
<td>4.47 (0.32)</td>
<td>8.731 (44)</td>
<td>&lt; .00</td>
</tr>
</tbody>
</table>

As the Table 3 depicts, there is a consistent pattern of statistically significant higher levels of improvement for the experimental group compared to the control group. This finding suggests that students considered that the FCM approach enhanced not only their satisfaction for the teaching practice, but also their own conceptualization of the ICT course and the learning process. Therefore, students in the experimental group were more confident in successfully engaging with the learning activities (Confidence dimension) and were also more intrigued by the manner in which the teaching process was delivered (Attention). Finally, the results from the Relevance dimension indicate that the students viewed the FCM-enhanced ICT course as being more relevant to their own interests.
Results regarding students’ level of engagement (RQ4)

The RQ4 aimed to investigate whether the exploitation of the FCM would lead to enhanced students’ engagement. Table 4 presents the results obtained from the independent samples t-test, based on the mean scores of engagement for each week of the study.

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>t(df)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group [N = 23]</td>
<td>3.61 (0.81)</td>
<td>3.230</td>
<td>&lt; .00</td>
</tr>
<tr>
<td>Experimental group [N = 23]</td>
<td>4.01 (0.31)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As the Table 4 depicts, there is a statistically significant difference between the two groups, in favor of the experimental group. It is evident that the FCM was beneficial not only for maintaining students’ level of engagement throughout the ICT course, but also for increasing it as the course delivery progressed. Furthermore, in order to investigate the level of engagement and its progress throughout the 8-week period, data from the within-group student clusters were analyzed and are presented in Figure 4 (control group) and Figure 5 (experimental group).

As the Figure 4 and Figure 5 depict, students in both groups showed a continuous increase in engagement, with a larger improvement margin in favor of the experimental group clusters. Moreover, both figures show a small decline pattern in students’ engagement levels between weeks #3 and #4 which is consistently reversed between weeks #5 and #6. These patterns can be explained in conjunction with the placements of the standardized assessment tests #2 and #3, which were delivered on week #2 and week #6 correspondingly.
Furthermore, regarding the experimental group, a significant finding is that the low performers cluster was most benefited (Figure 5). More specifically, the low performers showed an overall increase of 30.1% in their engagement level, compared to a 16.3% increase of the medium performers and a 7.8% of the high performers. Thus, it is argued that the FCM not only enhanced engagement for all students, but also provided the best benefits for the students that struggle in terms of performance. This finding is consistent with the findings from RQ1, namely that low performers were also most benefited from the FCM in terms of cognitive learning outcomes improvement.

**Discussion and future work**

The paper presented the design and implementation of an action research for the first attempt to evaluate the FCM in the context of K-12 ICT teaching. Based on a literature review on existing works in Higher Education, four research questions were defined. The added value of the action research is twofold, namely (a) it investigated the potential of FCM in the previously unexplored educational context of K-12 ICT teaching, and (b) it adopted a more holistic conceptualization of students’ learning experience, by combining existing approaches to model it in relevant works.

The overall insights from the action research are highly promising and consistent with the findings of similar works in the context of Higher Education. More specifically, the results indicated that the incorporation of the FCM in the teaching and learning process led to a statistically significant increase in the cognitive learning outcomes of students. This is a commonly reported benefit of the FCM in Higher Education ICT education, and this study provided evidence to extend this statement to the K-12 context, as well. Another added-value finding of this work was that the low performers were identified as the group that experienced the largest improvement. This can be attributed to the capacity of the students to receive formative feedback and scaffolding during face-to-face sessions, both from their teacher as well as from their classmates when engaging in collaborative activities.

Furthermore, the experimental group showed statistically significant increase in their motivation. This finding signifies that students’ satisfaction and interest in the ICT course was enhanced and, moreover, that students were able to link the learning process to their own interests and improve their sense of accomplishment. These findings are consistent with existing works, e.g., Gehringer and Peddycord (2013), and Dolgopolovas et al. (2014). Regarding students' engagement, the results provide useful new evidence considering that this aspect had previously received very little explicit attention (e.g., Herold et al., 2012). More specifically, the results indicated that the FCM provided two main benefits: (a) the students were significantly more engaged throughout the course with a continuously increasing trend, and (b) within-group cluster analysis in the experimental group revealed that the low performers had the largest percentage of improvement. These benefits (as well as the evidence regarding student motivation) can be largely attributed to the better exploitation of face-to-face sessions that the FCM promoted. As the analysis of the delivered types of teaching and learning activities indicated, the FCM allowed the practitioner to primarily focus on competence-building “hands-on” activities and formative feedback provision. Even though previous studies had also suggested this connection (e.g., Sureka et al., 2013; Davies et al., 2013) this work provided an explicit and transparent analysis of the specific types of learning activities employed in each group throughout the action research to support the statement.

Based on these empirical results, the key insights which can be elicited include:

- The FCM is appropriate for maximizing classroom time invested on collaborative, hands-on activities. Therefore, it can be exploited to promote engaging approaches to K-12 ICT teaching and learning (e.g., problem- and project- based approaches). Considering the reported benefits of such learning approaches to effectively promote core student skills and digital competences, such as creative coding and computational thinking (Sharplees et al., 2015), this promising potential of FCM should be further examined. Furthermore, the evidence related to increased frequency of student-student and student-teacher interaction during classroom-based sessions further supports this claim.

- The FCM is most beneficial for improving the learning experiences of low-performing students. This provides promising evidence for addressing a significant issue in K-12 ICT teaching, namely the commonly reported gender gap (Wang et al., 2015). It has been globally acknowledged that girls tend to have lower performance and engagement (or attitudes) towards ICT during K-12 Education, leading to a smaller percentage of female ICT professionals. In this context, the reported benefits of FCM could be further investigated in terms of addressing this global challenge.
Although our study provides evidence on students’ motivation, engagement and learning outcomes, there are also some limitations. First, the generalizability of the results must be carefully approached since the study was conducted within a specific context (e.g., educational level, course). Second, self-report methods (surveys and teacher journal) were used to measure the motivation and engagement variables, so some of the results might have a common method biases. However, the main focus of our study was to evaluate the FCM in K-12 ICT teaching through a more holistic lens, namely students’ motivation, engagement and cognitive learning outcomes. This reduces the common method bias as we used both attitudinal and cognitive variables. Nevertheless, additional research is needed to complement our findings based on deeper qualitative methods, such as in-depth interviews, interaction analyses and behavioral patterns.

Future work should include further research for explicitly evaluating other aspects of the potential of FCM towards enhancing students’ learning experiences in ICT (Giannakos et al., 2016), such as the impact on the affective status of students, the potential for addressing the widely acknowledged gender gap on performance and motivation towards ICT, as well as the capacity to foster creative computational thinking skills. Furthermore, more longitudinal approaches should be designed, so as to gain insights on the impact of the FCM on students’ learning experiences over a wider span of school time (more than one school year). Such longitudinal studies should also aim to capture, process and report additional qualitative evidence on how FCM can affect the learning process and students’ development of digital citizenship skills including computational thinking, collaboration skills and potentially minimizing the globally acknowledged gender gap. Finally, future work should also take into account the state-of-the-art in the emerging fields of Teaching and Learning Analytics (Sergis & Sampson, 2016; Sergis & Sampson, In Press). These fields investigate tools and methods to support teachers engage in evidence-based reflective re-design of their teaching practice based on the analysis of student educational data. The added-value of these emerging technologies is that they may also exploit the collection and processing of student data which can transcend action-based data and may include more fine-grained student characteristics, such as their emotions. Thus, they can offer a much richer evidence pool, which teachers can process and exploit when investigating and (re)designing their practice to enhance the provision of personalized learning experiences to their students.

References


Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. Education for information, 22(2), 63-75


Integrating the SOP² Model into the Flipped Classroom to Foster Cognitive Presence and Learning Achievements

Hsiu-Ling Chen¹ and Chiang-Yun Chang²

¹Department of Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // ²Department of Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // shirley@mail.ntust.edu.tw // M10211001@mail.ntust.edu.tw

Corresponding author

ABSTRACT

This study explored student teachers’ cognitive presence and learning achievements by integrating the SOP² Model in which self-study (S), online group discussion (O) and double-stage presentations (P²) were implemented in the flipped classroom. The research was conducted at a university in Taiwan with 31 student teachers. Pre- and post-worksheets measuring knowledge of educational issues were administered before and after group discussion. Quantitative content analysis and behavior sequential analysis were used to evaluate cognitive presence, while a paired-samples t-test analyzed learning achievement. The results showed that the participants had the highest proportion of “Integration,” but rarely reached “Resolution.” The participants’ achievements were greatly enhanced using the SOP² Model in terms of the scores of the pre- and post-worksheets. Moreover, the groups with a higher proportion of “Integration” (I) and “Resolution” (R) performed best in the post-worksheets and were also the most progressive groups. Both high- and low-rated groups had significant correlations between the “I” and “R” phases, with “I” → “R” in the low-rated groups but “R” → “I” in the high-rated groups. The instructional design of the SOP² Model can be a reference for future pedagogical implementations in the higher educational context.

Introduction

Many studies have mentioned the potential value of student-centered learning environments in which students engage actively in higher-order thinking tasks, particularly in the higher education context (Davies, Dean, & Ball, 2013; Berrett, 2012; Baepler, Walker, & Driessen, 2014). However, most instructors who are used to delivering lectures might not implement appropriate instructional methods to support student-centered learning (Kim, Kim, Khera, & Getman, 2014). Thus, creating a learner-centered environment still remains a challenge. According to Sahin, Cavlazoglu and Zeytuncu (2015), the flipped classroom model using online videos may increase students’ achievement. By previewing flipped materials, students can focus on their individual paces and needs, while in-class activities can be designed to promote higher-order thinking skills (Davies et al., 2013). Echoing flipped principles, our study offered flipped materials, both online video lectures and text-based articles, before class, to help the participants gain proper course knowledge. Research has revealed that computer-supported collaborative learning (CSCL) enables students to learn in groups of four to six by conducting online discussion based on designed tasks for problem-solving resolution (Walker, 2004). Besides, CSCL contexts with online technology support for group work enhances cognitive performance (Resta & Laferrière, 2007). Similarly, a CSCL environment serves as a useful tool for collaborative knowledge construction as a result of students’ discussion (Salovaara, 2005). Therefore, knowledge construction among groups could be promoted in CSCL contexts.

In addition to integrating CSCL contexts, researchers have reported that flipped learning enhances learning effectiveness through digital resources, group discussion, oral reports and teacher guidance (Li, Lou, Tseng, & Huang, 2013). It has also been revealed that the flipped classroom design with off-loaded contents, student discussion and presentations, and proper assessment, promotes concept understanding, knowledge application and critical thinking skills (McLaughlin et al., 2014). Moreover, Kong’s (2014) digital classroom design enabled learners to apply cognitive knowledge due to online group discussions, teacher feedback and worksheet questions. In fact, it is believed that instructional videos embedded with open-ended questions are effective in learning new content (Enfield, 2013). Thus, it has been suggested that the flipped classroom design adopting digital resources, group-based discussions, oral presentations, feedback sessions and worksheets designed with open-ended questions could foster cognitive development and promote knowledge learning. Accordingly, we propose the SOP² Model, where “S” represents the Self-study stage during which each participant previews the flipped materials and completes worksheets, “O” represents the Online group discussion stage during which each group conducts synchronous online discussion in a CSCL context and completes in-class worksheets, and the
According to the Community of Inquiry (CoI) framework, learning experiences are formed through the interaction of cognitive presence, teaching presence, and social presence (Garrison, Anderson, & Archer, 1999). Researchers have also pointed out that CoI can depict knowledge construction via online collaborative environments in terms of technology assistance, social interaction, and instructional procedures (Shea & Bidjerano, 2010). The designed principles of our proposed SOP² model correspond to three major factors: cognitive presence, teaching presence, and social presence from CoI. According to researchers, self-efficacy and metacognition are aspects of self-regulation, which is essential for online learning (Shea et al., 2012). In addition, researchers have suggested that higher-order learning in online environments can be evaluated through cognitive presence (Garrison, Anderson, & Archer, 2001). On the one hand, studying flipped materials prior to class requires much self-regulation and metacognitive learning skills. On the other, open-ended worksheets can assess participants’ higher-order learning results. Thus, the design of a “Self-study” stage can help explore the learners’ roles in terms of cognitive presence, self-regulation, and metacognition so as to understand how self-learning takes place in the online environment. Next, the “Online group discussion” stage corresponds to social presence in CoI. It is believed that online discussion can facilitate positive emotion, social interaction, and solidarity through emotional expression, open communication, and group cohesion (Shea & Bidjerano, 2010; Garrison et al., 1999). Thus, the design of synchronous online discussion aimed to help participants facilitate collaborative group work via social interaction and communication. Lastly, the “Double-presentation” stage involving group presentations, feedback sessions, and whole-class discussion reflects the spirit of “Teacher presence” in CoI. Garrison et al. (1999) revealed that teacher presence can be shown through three major principles, namely instructional management via organized instructional design, building understanding via personal or group sharing, and direct instruction via explanation or clarification by instructors. Similarly, the “Double-presentation” design consists of standardized instructional procedures with clear goals and learning contexts, group sharing of knowledge with feedback sessions, and whole-class discussion along with teacher-led summaries and explanations based on the group presentations.

Regarding the importance of the SOP² model, the online video-lectures offered during the self-study stage can improve students’ problem solving preparation and efficiency since the flipped materials could prompt them to revisit the videos to seek and answer relevant questions during problem-solving activities (Tawfiq & Lilly, 2015). Next, the synchronous online discussion stage was especially designed to allow participants to share and reflect on ideas without time and space limitations, as well as to learn from multiple perspectives to construct knowledge through interactive discourse (Lipponen, 2002). Similarly, Branon and Essex (2001) claimed that instant interaction in synchronous communication can facilitate knowledge sharing and feedback. Moreover, since the participants were required to conduct online discussion based on educational problem-solving contexts, it was discovered that synchronous discussion could enhance their decision-making and brainstorming skills (Branon & Essex, 2001) which are important abilities for problem-solving contexts. In addition, another study showed that learners’ problem-solving skills can be reinforced using synchronous computer-supported collaborative learning tools (Lazakidou & Retalis 2010). As for the presentation feedback and revision stages, a previous study suggested that whole-class discussions and oral peer assessments enable participants to increase their reflection on new situations as well as promoting self-assessment and awareness (Topping, 2009). Moreover, teacher feedback can assist learners in clarifying goals, enhancing commitment, and increasing their learning efforts (Hattie & Timperley, 2007).

Cognitive presence describes the process of community inquiry into a problem (Garrison, Anderson, & Archer, 2001). Garrison and colleagues (2001) measured cognitive presence in terms of the Practical Inquiry Model to evaluate the quality of online discourse. The PI Model consists of four cognitive phases, namely Triggering events, Exploration, Integration, and Resolution (Garrison & Anderson, 2003). Since this model is considered to be an effective tool for evaluating cognitive presence and higher-order learning (Schrire, 2006; Akyol & Garrison, 2011), in the present study, we have adopted it to analyze online discourse to gain an understanding of the participants’ level of cognitive presence. Akyol and Garrison’s (2011) study indicated that online collaborative development of cognitive presence is linked with actual learning grades. Thus, we also evaluated our participants’ course-specific knowledge based on their individual worksheet scores in order to understand how the implementation of the SOP² Model affected their learning achievements. The following research questions were thus formulated: (1) How does integrating the SOP² Model in the flipped classroom influence the student teachers’ cognitive presence? (2) How does integrating the SOP² Model in the flipped classroom influence the student teachers’ learning achievements? (3) In what way does cognitive presence relate to the learning achievements of the groups using our proposed model?
Method

Our study applied both quantitative and qualitative methods to investigate participants’ cognitive presence and learning performance. Online discourse analysis using the PI Model was used to explore cognitive presence, while individual worksheet scores were used for the evaluation of students’ learning achievements. In addition, student interviews were adopted as qualitative data to gain deeper understanding of our developed model. Integrating the SOP² Model in the flipped classroom was the independent variable. The first research question investigated the four categories of cognitive presence, the second explored learning achievements based on the scores of individual worksheets, while the third explored the relationship between cognitive presence and learning achievements of the groups. See Figure 1 for our research design structure.

![Figure 1. The research design structure](image)

Participants

This study was conducted using an eighteen-week course, “Introduction to Education” as the research context. The participants attended this course at a university in Taiwan in 2015. One instructor with 13 years of teaching experience, two teaching assistants, and 31 student teachers (23 undergraduates and 8 graduates) were involved. These student teachers were enrolled in a secondary teacher education program. The participants were assigned to groups of four, except for one group of three, with eight groups in total.

Instructional contexts and procedures

There were three cycles of SOP² sessions in the flipped classroom throughout the semester, and the format of each was the same. Our research design was based on a specific course plan. The course requirements were introduced in the first week. For the following two weeks, education theories were introduced during face-to-face classes. Meanwhile, grouping was conducted by mixing the participants with low, medium, and high critical thinking dispositions based on “The Inventory of Belief and Critical Thinking Disposition” (Yeh, 1999). Moreover, an icebreaker activity was conducted to increase familiarity among group members. Online discussion skills were introduced and techniques for operating the TMOF system were demonstrated in the fourth class.

Self-study stage

The participants had one week to preview the flipped materials at home before attending each class session. The materials, including online videos with PPTs, audio recording, and relevant articles, were uploaded to the online system. They were carefully designed and arranged by the instructor and the research team based on current educational issues. There were three issues in total which combined theories from psychology, sociology,
philosophy and administration of education, curriculum design and development, teaching techniques and strategies, class management, as well as youth counseling and guidance. See Table 1 for detailed descriptions of the three educational issues, the discussed topics, and relevant theories. Students could download the flipped materials through the online system to gain preliminary knowledge and understanding of the issues. Besides, each participant was required to explore the answers from the selected materials and complete an individual “Flipped Classroom Worksheet” with five open-ended questions. The participants had one week to read the materials and submit the individual worksheet through the online system.

<table>
<thead>
<tr>
<th>Educational issue</th>
<th>Issue 1: The Development of Twelve-year Compulsory Education</th>
<th>Issue 2: The Challenge of Technology Integration into Instruction</th>
<th>Issue 3: The Predicament of Neo Children in Taiwan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theories Included</td>
<td>• Psychology of education • Sociology of education • Teaching techniques and strategies • Course design and development • Administration of education</td>
<td>• Philosophy of education • Psychology of education • Sociology of education • Course design and development • Teaching techniques and strategies</td>
<td>• Sociology of education • Psychology of education • Youth counseling and guidance • Class management • Administration of education</td>
</tr>
</tbody>
</table>

**Online group discussion stage**

CSCL contexts were designed for the synchronous class time. The student teachers had their own computers and synchronously communicated with their group members through the TMOF system. Prior to online group discussion, the participants were required to prepare for the upcoming events. Also, face-to-face interaction was permitted before the CSCL contexts to establish ground rules for the online discussion. As a warm-up session, face-to-face interaction took place prior to the online discussion stage. Each group then conducted collaborative online discussion for 70 minutes via the online system, using chat forum, tree map, and other chatting tools. Each group presented, exchanged ideas, and constructed a shared tree map of their group discussion results. The instructor and two teaching assistants monitored and facilitated eight groups at a time by answering questions or solving technical problems. At the end of the group discussion, an individual “In-class Worksheet” with five open-ended questions was issued through the TMOF system. Each participant answered the assigned questions and referred to the information from the discussion forum or tree map outcomes. This stage lasted around 30 minutes.

**Double-presentation stage**

This stage came after each online stage ended. Each group organized their text-based discourse and tree map results based on the three assigned topics. Each of the eight groups presented their group report in the form of PPTs and demonstrated their 1st stage group presentation (6 minutes), focusing on sharing general knowledge. They received written and oral feedback from the instructor and peers before refining their 2nd stage group presentation (8 minutes), focusing on exploring the comprehensive knowledge in more depth. The feedback sessions enabled the participants to offer reflective and concrete comments based on the presentations. Individual worksheets and group reports were completed and submitted through the online system. Teacher announcements and e-mails were sent to remind the participants of upcoming events, class updates, and assignments. Figure 2 shows the cycle of instructional procedures in the SOP² Model.
Synchronous online discussions were collected for discourse analysis. Our coding instrument for evaluating the quality of cognitive presence was based on the PI model, which comprises four cognitive categories: (1) Triggering events (category T), (2) Exploration (category E), (3) Integration (category I), and (4) Resolution (category R). Each coded unit is “a meaningful message” (Kanuka, Rourke & Laflamme, 2007; Vaughan & Garrison, 2005). Each meaningful message is made up of a sentence or sentences with a definite idea or clear intention. For example, the context for educational issue one describes: “Cheng is a junior-high school student who usually falls asleep during class time and has difficulty concentrating on his studies.” One student from group 5 stated at the very beginning of the group discussion, “According to the context, he already lacked learning motivation.” This sentence shows a clear message with an attempt to present background information for the educational context. Online discourses were coded based on relative keywords associated with the definitions and indicators of the four cognitive categories proposed by Garrison and colleagues (2001, p. 17-20), while the other messages belong to others (category O). Each participant had their own code, where A to H represents the 8 groups and 1 to 4 represents the group members. Thus, the first member of group 1 was coded A1. Every meaningful message was counted once under each category. Quantitative content analysis was adopted to count the frequency for each category and the distribution over the four cognitive phases. Sequential behavior analysis was also adopted to explore whether a behavioral sequence of one discussion behavior followed by another was significant (Hou & Wu, 2011). Example discussion messages from the four cognitive phases are presented in Table 2.

**Figure 2. Cycle of instructional procedures in the SOP² Model**

**Instruments**

*Online discourse analysis and coding examples*

Synchronous online discussions were collected for discourse analysis. Our coding instrument for evaluating the quality of cognitive presence was based on the PI model, which comprises four cognitive categories: (1) Triggering events (category T), (2) Exploration (category E), (3) Integration (category I), and (4) Resolution (category R). Each coded unit is “a meaningful message” (Kanuka, Rourke & Laflamme, 2007; Vaughan & Garrison, 2005). Each meaningful message is made up of a sentence or sentences with a definite idea or clear intention. For example, the context for educational issue one describes: “Cheng is a junior-high school student who usually falls asleep during class time and has difficulty concentrating on his studies.” One student from group 5 stated at the very beginning of the group discussion, “According to the context, he already lacked learning motivation.” This sentence shows a clear message with an attempt to present background information for the educational context. Online discourses were coded based on relative keywords associated with the definitions and indicators of the four cognitive categories proposed by Garrison and colleagues (2001, p. 17-20), while the other messages belong to others (category O). Each participant had their own code, where A to H represents the 8 groups and 1 to 4 represents the group members. Thus, the first member of group 1 was coded A1. Every meaningful message was counted once under each category. Quantitative content analysis was adopted to count the frequency for each category and the distribution over the four cognitive phases. Sequential behavior analysis was also adopted to explore whether a behavioral sequence of one discussion behavior followed by another was significant (Hou & Wu, 2011). Example discussion messages from the four cognitive phases are presented in Table 2.
<table>
<thead>
<tr>
<th>Table 2. Example discussion messages of the four cognitive phases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Triggering events (category T)</strong></td>
</tr>
<tr>
<td><strong>Example discussion messages</strong></td>
</tr>
<tr>
<td>(Lead direction) Start from Q2: The cause and influence of digital gap</td>
</tr>
<tr>
<td>(Ask question) Does anybody know individualized learning?</td>
</tr>
<tr>
<td>(Lead new direction) I believe it is hard to separate Q1 into positive and negative aspects.</td>
</tr>
<tr>
<td>What is your opinion?</td>
</tr>
<tr>
<td>(Pose question) I have no idea how to analyze Q2 using theories from psychology of education ...</td>
</tr>
<tr>
<td>(Present background information) According to the context, he already lacked learning motivation...</td>
</tr>
<tr>
<td><strong>Explanation (category E)</strong></td>
</tr>
<tr>
<td><strong>Example discussion messages</strong></td>
</tr>
<tr>
<td><strong>(Contradictory viewpoints)</strong></td>
</tr>
<tr>
<td>Do we have to list out the reasons for students’ distractions?</td>
</tr>
<tr>
<td>But I divided into sensing problems and parent cooperation.</td>
</tr>
<tr>
<td><strong>(Personal descriptions)</strong></td>
</tr>
<tr>
<td>When I was in senior high school, we were not allowed to use cellphones.</td>
</tr>
<tr>
<td><strong>(Divergent viewpoints)</strong></td>
</tr>
<tr>
<td>Is multicultural education different from teaching adjustment?</td>
</tr>
<tr>
<td><strong>(Unsupporting addition)</strong></td>
</tr>
<tr>
<td>Even though there is Internet regulation, no one can monitor the students.</td>
</tr>
<tr>
<td><strong>(Personal opinions)</strong></td>
</tr>
<tr>
<td>Content design should motivate students’ learning interest and turn into internal learning motivations.</td>
</tr>
<tr>
<td><strong>Integration (category I)</strong></td>
</tr>
<tr>
<td><strong>Example discussion messages</strong></td>
</tr>
<tr>
<td><strong>(Connect ideas)</strong> I’m restating the educational differences between Finland and Taiwan; Finland is more focused on (1) individualized teaching and learning; (2) adjustable teaching approach; (3) adjustable assessments.</td>
</tr>
<tr>
<td><strong>(Build on the idea)</strong> Teachers of the same grade can exchange teaching approaches, communicate more frequently, inspire teaching strategies...Teachers should be required to take professional courses so they can work with teachers teaching different grades.</td>
</tr>
<tr>
<td><strong>(Offer relevant source)</strong></td>
</tr>
<tr>
<td>Based on the definition in the materials, “information literacy” means: Knowing when he/she needs the information and his/her ability to search, evaluate, and use the information.</td>
</tr>
<tr>
<td><strong>(Integrate relevant source)</strong></td>
</tr>
<tr>
<td>Volunteers can be motivated by their peers. Reference: <a href="http://www.epochtimes.com/b5/15/3/19/n4391081.htm">http://www.epochtimes.com/b5/15/3/19/n4391081.htm</a></td>
</tr>
<tr>
<td><strong>(Create possible solutions)</strong></td>
</tr>
<tr>
<td>Utilizing online contest system for students to make up questions and have others answer them, in order to enhance motivation and competitiveness.</td>
</tr>
<tr>
<td><strong>Resolution (category R)</strong></td>
</tr>
<tr>
<td><strong>Example discussion messages</strong></td>
</tr>
<tr>
<td><strong>(Comprehensive solution)</strong></td>
</tr>
<tr>
<td>Assessment can include school grades, students’ diary, worksheets, research and data collection, presentation skills, observations as well as interviews to evaluate students’ learning progress.</td>
</tr>
<tr>
<td><strong>(Evaluate the solution)</strong></td>
</tr>
<tr>
<td>Helping students find their interest is a good point, but setting higher goals for them needs further evaluation of their learning ability.</td>
</tr>
<tr>
<td><strong>(Test the solution)</strong></td>
</tr>
<tr>
<td>It is good to explore his strengths and give him a sense of achievement; meanwhile, we need to acknowledge his peer relationship. It is win-win if handled well, otherwise it is lose-lose.</td>
</tr>
<tr>
<td><strong>(Alternative solution)</strong></td>
</tr>
<tr>
<td>This activity can also be applied to Q5, which concerns how to lessen bullying and cliques. This approach helped them learn to accept different ideas of individuality.</td>
</tr>
<tr>
<td><strong>Other (category O)</strong></td>
</tr>
<tr>
<td><strong>Example discussion messages</strong></td>
</tr>
<tr>
<td><strong>(Monitor discussion situation)</strong></td>
</tr>
<tr>
<td>Any other opinions? Or we can move on to the next</td>
</tr>
<tr>
<td><strong>(Remind time)</strong></td>
</tr>
<tr>
<td>We had better hurry up. We had two more</td>
</tr>
<tr>
<td><strong>(Distribute tasks)</strong></td>
</tr>
<tr>
<td>We pick one person to draw the tree map and other members can express their opinions</td>
</tr>
<tr>
<td><strong>(Express agreement)</strong></td>
</tr>
<tr>
<td>You are great! I’ll give you a “like”.</td>
</tr>
<tr>
<td><strong>(Greeting)</strong></td>
</tr>
<tr>
<td>Hello! Good afternoon everyone!</td>
</tr>
</tbody>
</table>
Individual worksheets

The “Flipped Classroom Worksheet” was considered as the “Pre-worksheet,” while the “In-class worksheet” was the “Post-worksheet.” The worksheets and rubrics were designed by our research team and scrutinized by an expert with more than ten years of teaching experience in a teacher education program. The pre- and post-worksheets consisted of five open-ended questions regarding the educational topics and related theories with the purpose of measuring the students’ learning achievements. The questions are literally the same with slight structural differences between the two worksheets (see the Flipped Classroom Worksheet example in Figure 3 and the In-class Worksheet example in Figure 4). The rubric for grading individual worksheets is provided in Table 3. To validate the worksheet contents, preliminary analysis of the correlation between the pre- and post-worksheets was performed. Table 4 shows that pre-worksheets regarding issue 1, 2, and 3 had significant correlations with their post-worksheets, showing $p < .05$, $p < .01$, and $p < .001$, respectively.

Table 3. The rubric for individual worksheets

<table>
<thead>
<tr>
<th>Rubrics</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Each individual worksheet contains five open-ended questions</td>
<td>Maximum 20 points for each question</td>
</tr>
<tr>
<td>(2) Offer logical and corresponding answers to the questions</td>
<td>10 points for each question</td>
</tr>
<tr>
<td>(3) List the answers by bullet points: higher scores for more listed items</td>
<td>a. 1 point for each listed item</td>
</tr>
<tr>
<td>(4) None of these three conditions gets any points:</td>
<td>b. 1 extra point for each listed item with rich contents</td>
</tr>
<tr>
<td>a. Directly copy and paste all the information</td>
<td>c. Simply answer “Yes” or “No”</td>
</tr>
</tbody>
</table>

Q1. (1) Please list out at least three points after watching the video on educational views from Finland; (2) Please point out a concept or a sentence you are most impressed with in the video regarding learning motivation. Please explain how you can increase students’ learning motivation.

Q2. Please explain the impacts of social services and elaborate the value and meaning of social services.

Q3. (1) According to the video clip, please explain the features of a learning community; (2) Please elaborate how teachers can implement differentiated instruction to decrease students’ proficiency gap.

Q4. What principles should be included in course design? Please explain in detail.

Q5. What are the difficulties with the “Twelve-year Compulsory Education” policy? Please list out relevant solutions for the proposed difficulties.

Figure 3. Flipped classroom worksheet example

| Q2. What are the influences of social services? Please discuss the value and meaning of social services. |
| Q3. Due to diverse social structures, there is an increasing learning gap among students. Please elaborate instructional methods for differentiated learning. |
| Q4. Please elaborate how to arrange and design courses so as to meet students’ learning needs. |
| Q5. In your opinion, what are the problems triggered by the “Twelve-year Compulsory Education” policy? Please explain possible solutions for the problems. |

Figure 4. In-class worksheet example

Table 4. Correlation between pre- & post-worksheets

<table>
<thead>
<tr>
<th>Individual worksheet</th>
<th>N</th>
<th>Correlation</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue1</td>
<td>31</td>
<td>.383</td>
<td>.033&quot;</td>
</tr>
<tr>
<td>Issue2</td>
<td>30</td>
<td>.479</td>
<td>.007**</td>
</tr>
<tr>
<td>Issue3</td>
<td>31</td>
<td>.589</td>
<td>.000***</td>
</tr>
</tbody>
</table>

Note. "p < .05, **p < .01, ***p < .001.

Each participant’s two achievement scores from the pre- and post-worksheets were evaluated by dependent $t$-test to investigate learning achievement. Two raters reached a compromise concerning the rubrics before actual grading. Around 20 worksheets were randomly picked for an inter-rater reliability test. The correlation coefficient agreement between the two raters reached .965 ($p = .000$), indicating high reliability. To answer in
what ways cognitive presence relates to learning achievements among groups, we compared the results of cognitive presence and worksheet performance of the eight groups. Pearson Product Moment correlation was operationalized to analyze how the achievement scores were associated with the cognitive presence of the groups.

Results and discussion

Cognitive presence

A total of 1,134 units of “meaningful message” were coded. Of these, 114 (10%) were randomly selected to conduct Kappa coefficient analysis. The kappa value was .806 which reached the .000 level of significance between the two raters. Regarding the frequency and distribution of the cognitive categories, overall class discussions demonstrated that the largest proportion is category “E” (n = 408; 36.0%), the second largest is category “F” (n = 257; 22.7%), the third is category “T” (n = 235; 20.7%), the fourth is category “O” (n = 199; 17.5%), and the lowest proportion is Category “R” (n = 35; 3.1%) (see Table 5).

Comparing the frequency and distribution of the eight groups, five (G1, 2, 4, 5, 8) had the highest proportion of category “E,” while the other three fell into different categories. Four groups (G1, 3, 7, 8) had the second highest proportion of category “I.” The proportion of the category of “Resolution” was relatively low, with no more than 8.1% among the eight groups; one group (G5) had no coded discussion in the “R” phase. The distribution of category “O” varied among the eight groups. The highest percentage was 43%, whereas the lowest was 5%. The percentage of category “T” also showed little difference among groups, with the highest percentage nearly 32%, and the lowest around 4%. See the distribution of cognitive categories for the whole class in Figure 5.

The results indicate that both the whole class and individual groups had the highest frequency of exploring or sharing different opinions and experiences. The “Exploration” phase dominating cognitive presence is consistent with previous studies (Kanuka, Rourke, & Laflamme, 2007; Vaughan & Garrison, 2005; Garrison et al., 2001). The second most frequent discussion behavior for the whole class and among groups was “Integration,” during which the students attempted to connect or extend other members’ ideas and create possible solutions. Although the results showed that “Integration” is the second most active phase, compared to previous research (Kanuka et al., 2007; Vaughan & Garrison, 2005; Meyer, 2003) which showed a drastic reduction in this phase, the distribution rate of “Integration” in our study is rather high among the four categories of cognitive presence. Additionally, according to our student interviews, numerous participants revealed that our instructional approach enabled them to reach the “Integration” level by connecting, synthesizing solutions and further constructing educational knowledge.

Lastly, the “Resolution” phase rarely occurred, having the lowest proportion during synchronous group discussions both in the whole class and among groups. A number of students claimed that the online discussion duration was too short for thorough discussions. According to Akyol and Garrison (2011), lack of time hinders the “Resolution” phase since students do not have sufficient time to share their application results. Moreover, based on the student interviews, owing to their lack of teaching experience, the participants could only discuss the topics in terms of the theories introduced in the flipped materials and they had difficulty offering concrete resolutions. In fact, we found that 11 of the 31 participants, around 1/3, reported that they did not have any hands-on teaching experience.

Table 5. Cognitive category frequency distribution for the whole class

<table>
<thead>
<tr>
<th>Group/Category</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
<th>G7</th>
<th>G8</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>T N</td>
<td>17</td>
<td>43</td>
<td>3</td>
<td>29</td>
<td>90</td>
<td>15</td>
<td>27</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>12.2%</td>
<td>25.6%</td>
<td>4.1%</td>
<td>18.8%</td>
<td>31.8%</td>
<td>13.6%</td>
<td>26.2%</td>
<td>10.7%</td>
<td></td>
</tr>
<tr>
<td>E N</td>
<td>70</td>
<td>46</td>
<td>15</td>
<td>55</td>
<td>105</td>
<td>37</td>
<td>20</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>50.4%</td>
<td>27.4%</td>
<td>20.3%</td>
<td>35.7%</td>
<td>37.1%</td>
<td>33.6%</td>
<td>19.4%</td>
<td>58.3%</td>
<td></td>
</tr>
<tr>
<td>I N</td>
<td>42</td>
<td>29</td>
<td>18</td>
<td>31</td>
<td>48</td>
<td>40</td>
<td>26</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>30.2%</td>
<td>17.3%</td>
<td>24.3%</td>
<td>20.1%</td>
<td>17.0%</td>
<td>36.4%</td>
<td>25.2%</td>
<td>22.3%</td>
<td></td>
</tr>
<tr>
<td>R N</td>
<td>5</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>2.2%</td>
<td>4.8%</td>
<td>8.1%</td>
<td>3.2%</td>
<td>0%</td>
<td>6.4%</td>
<td>4.9%</td>
<td>1.0%</td>
<td></td>
</tr>
<tr>
<td>O N</td>
<td>7</td>
<td>42</td>
<td>32</td>
<td>34</td>
<td>40</td>
<td>11</td>
<td>25</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

281
In fact, according to Garrison and colleagues (2001), reaching the “Resolution” phase requires clear expectations and opportunities so as to apply created knowledge. In other words, since some of the student teachers had no teaching experience, their discussions remained at the “Integration” level and failed to connect with the actual teaching environment, let alone to apply what they had learned to the teaching job.

Learning achievements

The Cronbach’s alpha value reached high reliability among the three sets of pre- and post-worksheets, reaching .858 and .843, respectively. A total of 24 worksheets were randomly picked to evaluate the inter-rater reliability. The correlation coefficients reached .965 (p = .000), indicating high reliability. Based on the three respective educational issues, as shown in Table 6, the participants made significant improvement from the pre- to the post-worksheets for educational issues 1, 2, and 3 with p < .01, p < .001, and p < .01, respectively.

Table 6. Paired-sample t-test between pre- and post-worksheets

<table>
<thead>
<tr>
<th>Individual worksheet</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flipped classroom (P)</td>
<td>31</td>
<td>42.61</td>
<td>16.78</td>
<td>-3.887</td>
<td>.001**</td>
</tr>
<tr>
<td>In-class (P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flipped classroom (P)</td>
<td>30</td>
<td>42.87</td>
<td>16.54</td>
<td>-3.567</td>
<td>.001**</td>
</tr>
<tr>
<td>In-class (P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. **p < .01, ***p < .001.

Paired-samples t-tests were performed to investigate the overall learning achievements by integrating online group discussion in the flipped classroom context. Table 7 shows that the participants’ overall learning achievements made significant progress from the pre- to post-worksheets (p < .001).

Table 7. Paired-sample t-test between pre- & post-worksheets

<table>
<thead>
<tr>
<th>Individual Worksheet</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issue 1.2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flipped classroom (P)</td>
<td>30</td>
<td>129.50</td>
<td>16.77</td>
<td>-4.821</td>
<td>.000***</td>
</tr>
<tr>
<td>Online (P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ***p < .001.

The paired-samples t-test results showed that the student teachers improved significantly in terms of their learning achievements as a result of integrating the SOP2 Model. According to the student interviews, the online...
discussions helped facilitate the exchange of multiple opinions as well as connect ideas from different members’ perspectives, so the completeness and broadness of the discussions were reinforced. Additionally, many students revealed that the online discussion activities enhanced the organization of the group discourse because proposed opinions and suggestions could be integrated into complete answers for the discussed topics. In this case, the student teachers’ learning achievements could be greatly improved when integrating the SOP² Model into the flipped classroom. The results are consistent with the previous study that group collaborative activities in web-based courses can significantly improve grades (Benbunan-Fich & Arbaugh, 2006).

Participant (D1): Group discussion helped me understand the opinions of other group members. I had opportunities to view the issue from different perspectives...the discussion results became broader and more complete.

Participant (D3): Basically every member’s opinions and suggestions were induced, and the discussion results combined the ideas from all the other members.

Cognitive presence and learning achievements

Further comparison of learning achievements was conducted among the eight groups. Based on the post-worksheet performance and the discrepancy score, Table 8 shows that the three highest rated groups were groups 1, 3, and 6, categorized as the “High-achievement groups,” while the three lowest rated groups, groups 4, 5, and 8, were categorized as the “Low-achievement groups.”

<table>
<thead>
<tr>
<th>Group</th>
<th>Flipped classroom (3 Pre-worksheets in total)</th>
<th>In-class (3 Post-worksheets in total)</th>
<th>Score of discrepancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>138.00</td>
<td>161.33</td>
<td>23.00</td>
</tr>
<tr>
<td>3</td>
<td>132.25</td>
<td>156.25</td>
<td>24.00</td>
</tr>
<tr>
<td>6</td>
<td>129.50</td>
<td>155.00</td>
<td>25.50</td>
</tr>
<tr>
<td>2</td>
<td>126.50</td>
<td>149.75</td>
<td>23.25</td>
</tr>
<tr>
<td>7</td>
<td>129.50</td>
<td>145.75</td>
<td>16.25</td>
</tr>
<tr>
<td>4</td>
<td>133.00</td>
<td>129.33</td>
<td>-3.66</td>
</tr>
<tr>
<td>8</td>
<td>125.50</td>
<td>128.50</td>
<td>3.00</td>
</tr>
<tr>
<td>5</td>
<td>124.50</td>
<td>121.00</td>
<td>-3.50</td>
</tr>
<tr>
<td>Total</td>
<td>129.50</td>
<td>143.23</td>
<td>13.73</td>
</tr>
</tbody>
</table>

In terms of the correlation between the percentage (%) of cognitive categories and learning achievements (see Table 9), there was no significant correlation between categories “T,” “E,” and “O” and the learning achievements of the eight groups. However, it was notable that category “I” correlated significantly with the post-worksheets (Pearson = .655; p = .039), indicating that the groups with a higher proportion of category “I” had better performance on the post-worksheets.

Table 9 also shows that the “Resolution” phase had significant correlation with the post-worksheets (Pearson = .695; p = .028) for the eight groups. We discovered that the higher the proportion of “Resolution”, the better performance on the post-worksheets. In addition, there was also a significant correlation between category “R” and the score of discrepancy (Pearson = .727; p = .02). That is, the groups with a higher rate of “Resolution” were also the most progressive groups. Quantitative content analysis was performed to further compare the differences between the high- and low-achievement groups. Table 10 shows the frequency distribution of cognitive categories for the high- and low-achievement groups. The high-achievement groups had 1.5 times (31%) the proportion of category “I” than the low-rated groups (19%). That is to say, the high-achievement groups had more frequent integrated messages and were more likely to develop concrete solutions for the discussed issues. Although the low-rated groups had around double the “Triggering events” of the high-rated groups, codes of “Resolution” in the high-rated groups occurred 2.5 times (frequency = 16) more than in the low-rated groups (frequency = 6). In other words, the high-rated groups were more inclined to create comprehensive resolutions and further apply possible solutions. For the category of “Exploration,” the high- and low-rated groups reached similar distributions, with 38% and 41% respectively. Meanwhile, the two groups had the same proportion (15%) of category “O.”
The sequential analysis was also conducted to understand the differences in the behavioral patterns of the high- and low-rated groups. The sequential analysis of coded discussions in the high- and low-achievement groups is shown in the adjusted residuals table (see Table 11), in which each row represents a starting behavior and each column represents a follow-up behavior (Hou & Wu, 2011). If a Z-score is greater than 1.96, the sequence of a row and a column is statistically significant \((p < .05)\) (Bakeman & Gottman, 1997). All the significant sequences in the high- and low-achievement groups are shown as behavioral transition diagrams in Figure 6.

The sequential analysis explores the differences in the behavioral patterns of knowledge construction of the high- and low-rated groups. Figure 6 indicates that their sequential behavior patterns were similar. Table 11 shows that both groups had six significant sequences with Z-scores greater than 1.96. In addition, both groups had significant sequences within its own dimension for categories “T,” “E,” “I,” “R,” and “O” (i.e., “T” \(\rightarrow\) “T”; “E” \(\rightarrow\) “E”; “I” \(\rightarrow\) “I”; “R” \(\rightarrow\) “R”; “O” \(\rightarrow\) “O”). While categories “T,” “E” and “O” were all independent from the other four dimensions, categories “I” and “R” were significantly correlated with each other in both groups. However, the high- and low-rated groups had different directional sequences between categories “I” “and “R.” The slightly different sequences were “R” \(\rightarrow\) “I” for the high-achievement groups, but “I” \(\rightarrow\) “R” for the low-achievement groups. This finding indicates that while the low-rated groups gradually formed comprehensive resolutions from the integration phase, the high-rated groups paid more attention to integrating the concrete resolutions once they were raised.

The proportion of the “Integration” and “Resolution” categories correlated significantly with the participants’ learning achievements. The groups with a higher proportion of the “Integration” phase had better post-worksheet performance. The groups with a higher rate of the “Resolution” phase also did better on the post-worksheets and were the most progressive groups based on the discrepancy score. Regarding quantitative content analysis, the high-rated groups outperformed the low-rated groups in the “Integration” and “Resolution” phases during knowledge construction, while triggering events, exploration of knowledge, and off-topic discussions had no direct impact on learning achievements. The “Integration” and “Resolution” phases are considered to be higher levels of cognitive presence (Garrison et al., 2001), and higher cognitive presence in collaborative online discussion environments is linked with actual learning grades (Akyol & Garrison, 2011). Thus, our results are consistent with previous findings, which showed that higher-rated groups with larger distribution in the “Integration” and “Resolution” stages have better learning achievements. Based on the student interviews, the high-rated groups (G3&6) were already well prepared with relevant information and answers before the class sessions. Thus, it can be inferred that the high-achievement groups were more likely to achieve the “Integration”

### Table 9. Correlation between cognitive presence and learning achievements

<table>
<thead>
<tr>
<th>Category (%)</th>
<th>Pre-worksheets</th>
<th>Post-worksheets</th>
<th>Score of discrepancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>T (%) Pearson correlation</td>
<td>499</td>
<td>483</td>
<td>393</td>
</tr>
<tr>
<td>p</td>
<td>.104</td>
<td>.113</td>
<td>.168</td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>E (%) Pearson correlation</td>
<td>.26</td>
<td>.290</td>
<td>.356</td>
</tr>
<tr>
<td>p</td>
<td>.475</td>
<td>.243</td>
<td>.193</td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>I (%) Pearson correlation</td>
<td>.489</td>
<td>.655*</td>
<td>.597</td>
</tr>
<tr>
<td>p</td>
<td>.110</td>
<td>.039</td>
<td>.059</td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>R (%) Pearson correlation</td>
<td>.270</td>
<td>.695*</td>
<td>.727</td>
</tr>
<tr>
<td>p</td>
<td>.259</td>
<td>.028</td>
<td>.020</td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>O (%) Pearson correlation</td>
<td>.025</td>
<td>.182</td>
<td>.210</td>
</tr>
<tr>
<td>p</td>
<td>.476</td>
<td>.333</td>
<td>.309</td>
</tr>
<tr>
<td>N</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

**Note:** \(p < .05\).

### Table 10. Cognitive distribution of high- and low-achievement groups

<table>
<thead>
<tr>
<th>Category</th>
<th>T</th>
<th>E</th>
<th>I</th>
<th>R</th>
<th>O</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-achievement groups (n = 3)</td>
<td>Frequency</td>
<td>35</td>
<td>122</td>
<td>100</td>
<td>16</td>
<td>50</td>
</tr>
<tr>
<td>Percentage</td>
<td>11%</td>
<td>38%</td>
<td>31%</td>
<td>5%</td>
<td>15%</td>
<td>100%</td>
</tr>
<tr>
<td>Low-achievement</td>
<td>Frequency</td>
<td>130</td>
<td>220</td>
<td>102</td>
<td>6</td>
<td>82</td>
</tr>
<tr>
<td>Percentage</td>
<td>24%</td>
<td>41%</td>
<td>19%</td>
<td>1%</td>
<td>15%</td>
<td>100%</td>
</tr>
</tbody>
</table>
and “Resolution” phases during the group discussions. In addition, all members were already assigned specific tasks before the online discussion.

Table 11. Sequential behaviors of high- and low-achievement groups

<table>
<thead>
<tr>
<th></th>
<th>Z-score</th>
<th>T</th>
<th>E</th>
<th>I</th>
<th>R</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-achievement groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>3.17</td>
<td>-1.48</td>
<td>-1.81</td>
<td>1.91</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>-2.97</td>
<td>5.33</td>
<td>0.71</td>
<td>-3.22</td>
<td>-3.64</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>-0.16</td>
<td>-1.09</td>
<td>2.45</td>
<td>0.06</td>
<td>-1.60</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>-0.58</td>
<td>-2.17</td>
<td>2.21</td>
<td>3.77</td>
<td>-1.72</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>1.84</td>
<td>-3.19</td>
<td>-3.86</td>
<td>0.35</td>
<td>7.55</td>
<td></td>
</tr>
<tr>
<td>Low-achievement groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>3.95</td>
<td>-0.46</td>
<td>-3.00</td>
<td>-1.39</td>
<td>-0.39</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>-4.08</td>
<td>6.94</td>
<td>-0.40</td>
<td>-2.05</td>
<td>-3.64</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>0.71</td>
<td>-4.13</td>
<td>4.73</td>
<td>3.02</td>
<td>-1.25</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>-1.26</td>
<td>-0.96</td>
<td>1.20</td>
<td>8.31</td>
<td>-0.94</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.44</td>
<td>-4.21</td>
<td>-1.35</td>
<td>-1.04</td>
<td>7.09</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05.

Figure 6. Behavioral diagrams of high and low-achievement groups

With respect to the sequential analysis of behaviors, the results showed no sequential correlation among the categories “T,” “E,” and “O” in either group, while two groups demonstrated significant correlations between the integration and resolution phases, whose sequential behavior could be an essential factor in the collaborative group discussion. The slight disparity lies in the different directional sequence between categories “I” and “R”, which resulted in high and low achievements.

Researcher: How did you keep your discussion focused?
Participant (F2): All the members were active and brought their own information to the discussion. I personally prepared a lot of stuff for discussion...we took turns sharing the ideas we had prepared in advance.
Researcher: Do you mean you had all the answers and information written down for the discussion?
Participant (C2): Yes. Every member did so and we had already organized our own answers. In fact, we even finished distributing tasks before the discussion activity.

By observing the online discourse patterns of the high-rated groups (G3&6), it is indicated that they integrated or organized the comprehensive resolutions all over again after each one was developed. Thus, the pattern showed a significant correlation from the “R” to “I” categories. In contrast, we discovered that the low-rated groups (G5&8) posted all the information and ideas on the system without careful organization, which is demonstrated in the following comments from the excerpts of interviews with the low-rated groups.
Since higher cognitive presence such as in the “Integration” and “Resolution” phases is connected with actual learning performance in collaborative online discussion environments (Akyol & Garrison, 2011; Garrison et al., 2001), judging from our study results, the high-rated groups had a higher proportion of activity in the “Integration” and “Resolution” stages, and thus they were more likely to gain better learning grades. A further analysis on the student interviews indicated that the high-rated groups had better preparation of available resources and possible answers to reach resolutions for the educational contexts. In addition, the student interviews revealed that the high-rated groups conducted task distribution before the discussion activities. It has been revealed that role assignment could increase peer interactions and contribute to a higher level of cognitive presence and knowledge construction (Bliss & Lawrence, 2009; DeWever et al., 2010; Gašević et al., 2015)

Therefore, due to the limited discussion time, it was more possible for the high-rated groups to reach the “Integration” and “Resolution” phases during group discussions. In addition, according to Mayordomo and Onrubia (2015), group organization and coordination are factors that distinguish high and low group performance during collaborative knowledge construction. Thus, by observing the online discourse, the sequential behavioral patterns suggested that the high-achievement groups, with a significant sequence from “Resolution” to “Integration,” tended to summarize or organize the proposed resolutions all over again at the end of the group discussion, while the low-achievement groups simply provided information and unorganized ideas without any final re-organization.

**Student perceptions on the SOP2 model**

Self-studying flipped materials increased motivation and triggered problem-solving (Triggering Event): Most participants claimed that the flipped materials triggered their interest in the discussion topics and drove them to search for relevant information from other resources in order to clarify questions related to the problem-solving context, if they failed to understand the ideas from the flipped materials.

**Researcher:** Do you mean you searched for additional resources?

**Participant (A1):** Yes, I believe the flipped materials served as a triggering point.

**Participant (F1):** After reading the flipped materials, I knew what other information was needed to solve the problems in the educational contexts.

Online group discussion facilitated exploration of different opinions and knowledge exchange (Exploration): Many students claimed that online group discussion assisted them in exploring and exchanging different perspectives and opinions with other group members, and thus the depth and completeness of their group discussions were reinforced.

**Researcher:** Do you think group discussion is an important instructional design?

**Participant (D1):** I believe group discussion helped me understand opinions from other group members. I had opportunities to view the issue from different perspectives... the discussion became broader and more complete.

Online discussion and the worksheet design reinforced idea connection and knowledge construction (Integration): Numerous students revealed that the online discussion helped connect and integrate diverse opinions, draw conclusions for the discussion outcomes, and achieve knowledge construction for the educational contexts. In addition, many other students revealed that the flipped classroom worksheet helped them organize the main points in the flipped materials, while the in-class online worksheets helped them integrate the group discussion outcomes and draw complete conclusions. They also mentioned that the in-class worksheets, which were closely connected with educational contexts, enabled them to clarify and evaluate the main points from the discussion results.

**Researcher:** Did group discussion help you complete the online individual worksheets?

**Participant (A2):** The flipped classroom worksheets only included my own thinking, while the in-class online worksheets included opinions and suggestions from the group discussion, and thus the contents of the in-class online worksheets were the group members’ integrated ideas.

**Participant (B4):** Yes. A lot of information could be connected as our own reference. The discussion results came from all the members' integrated ideas.
The double-presentation and the SOP² Model helped apply educational knowledge in the actual teaching and resulted in self-reflection (Resolution): The participants acquired new perspectives from many case studies and pedagogical practices shared in group presentations. Those students with actual teaching experience claimed that group demonstrations enabled them to reflect on previous teaching contexts, while those without teaching experience could catch a glimpse of authentic teaching environments. Additionally, many of the student teachers revealed that the group presentations not only broadened their horizons in terms of content knowledge, but also changed their original ideas about giving presentations. That is, they referred to other groups to adjust their own presentations.

| Researcher: Did the group presentations help inspire you to learn new knowledge? |
| Participant (E1): Yes, one specific group was very creative and grabbed everyone’s attention, making us think about how we could demo our presentations next time. We learned from other groups’ presentations about how to present. |
| Participant (G1): Group presentations broadened my horizons, especially the practical cases in actual teaching jobs, so I got to take a look at what happens in the real teaching environment. |

Quite a few students revealed that thanks to the SOP² Model, they could understand the educational issues thoroughly, and that the course design was helpful for connecting previous teaching experiences. Some students reflected on how to deal with problems if they were to experience similar educational contexts. This indicates that they were able to apply the learnt knowledge to the actual teaching. Furthermore, many students pointed out that the SOP² Model, especially the flipped materials, individual worksheet design, group presentations, and feedback sessions promoted logical thinking and self-reflection.

| Researcher: Did the course design help connect with actual teaching jobs? |
| Participant (D1): I imagined what I would do if the same scenario occurred in my classroom. I also reflected on how I implemented the English teaching approach when I taught in the cram school and what the better solutions were. |
| Researcher: Would you try to apply what you have learned in your teaching jobs? |
| Participant (C2): Certainly. Because I learned not only theories but also many real cases which I could apply in the teaching contexts right away. |
| Researcher: Did the course design help facilitate self-reflection? |
| Participant (A2): Yes. I thought about the purpose of the worksheet questions and how to present the answers logically...and also how to present the tree map in the group presentations. Feedback from the teacher and peers also made me reflect on what I had learned and what should be improved after the course. |
| Participant (B3): I reflected on the teacher feedback and good presentations whose contents were new to me. I wondered why I had never thought of that before. |

Conclusions

Our study, based on the Community of Inquiry (CoI) framework (Garrison, Anderson & Archer, 1999), developed a practical flipped classroom model (the SOP² Model), which was aimed to reinforce students’ learning experiences through the interaction of cognitive presence, teaching presence, and social presence. Most previous CoI studies focused on online courses and CSCL contexts, but the present study further extended the realm to the flipped classroom context and demonstrated how the integrated SOP² Model in the flipped classroom could influence the student teachers’ cognitive presence and learning achievements. Compared to previous CoI studies (Lambert & Fisher, 2013; Szeto, 2015), the present study extended the depth of analysis of cognitive presence (in all four dimensions: trigger event, exploration, integration, and resolution) in the flipped classroom context, with results showing that groups with a higher proportion of “Integration”(I) and “Resolution”(R) performed better on the post-worksheets and were also the most progressive groups. The major conclusions of the study are discussed as follows.

The impact of the SOP² model on learning achievements and cognitive presence

The paired-samples t-test results showed that the student teachers had much better performance on the post-worksheets. That is, it is indicated that improvements might have been made due to the flipped classroom integrating synchronous online discussion activities. Our instructional context increased learning effectiveness in terms of grades. As for cognitive presence, the student teachers had the largest proportion of category “E”. During the “Exploration” phase students exchanged and shared divergent ideas, opinions, and experiences. The
second largest proportion was the “Integration” phase, in which the distribution rate of cognitive presence was rather high compared to previous studies (Kanuka et al., 2007; Meyer, 2003), implying that the participants were able to connect and integrate various ideas as well as form possible solutions for the discussed topics. However, the student teachers rarely reached the “Resolution” phase due to limited discussion time and a lack of teaching experience, which stopped expectations or opportunities for the application of the learnt knowledge. More time for discussions and for bringing more practical teaching cases into class might be beneficial for the reinforcement of students’ cognitive presence.

Cognitive presence and learning achievements among groups

While the distribution of categories “T,” “E” and “O” had no direct correlation with learning achievements, categories “I” and “R” correlated significantly with the post-worksheets and the score of discrepancy among groups. Results showed that the groups with a higher proportion of categories “I” and “R” had better performance on the post-worksheets. Additionally, the groups with a higher rate of the “R” phase were also the most progressive groups. Furthermore, the high-rated groups outperformed the low-rated groups in the “I” phase, with a higher frequency of organized solutions, and also in the “R” phase, with a higher possibility of offering solutions. Thus, higher cognitive levels including the “Integration” and “Resolution” phases were associated with better learning achievements in the collaborative online discussion environment (Akyol & Garrison, 2011; Garrison et al., 2001).

In terms of sequential behavioral patterns, both the high- and low-rated groups had significant sequences within their own dimension in the four cognitive categories and a significant correlation between the “Integration” and “Resolution” phases, which could be a critical factor in collaborative group discussion. It was also worth mentioning that there is only a slight difference between the high- and low-rated groups. The discussion patterns indicated that the high-rated groups listed out all the proposed resolutions in an integrated way at the end of the group discussion. Thus, the difference in directional correlation between the two phases resulted in disparity of high and low learning achievements. This finding indicates that while the low-rated groups gradually formed comprehensive resolutions from the integration phase, the high-rated groups paid more attention to integrating the concrete resolutions once they were raised.

Perceptions of the SOP² model for cognitive presence

The participants believed that the self-studying flipped materials motivated learning interest and problem-solving in educational contexts (Triggering event), while the online group discussion helped them explore knowledge and exchange diverse information (Exploration). Furthermore, the online discussion activities and worksheet designs helped integrate proposed solutions and construct knowledge related to educational contexts (Integration). Moreover, the two-stage presentations changed the students’ original mindset regarding giving group presentations. Lastly, the SOP² Model helped associate the knowledge with the actual teaching environment and helped the students engage in personal reflection (Resolution).

Pedagogical implications

Offering practical pedagogy and educational knowledge to prospective teachers

The importance of incorporating information technologies into education has been an important topic since the beginning of in the 21st century, and higher-order thinking skills have since been considered essential abilities (Salpeter, 2003). Our instructional approach enables the student teachers to experience technology-supported and higher-order learning activities as well as the global trends of ICT contexts, including autonomous learning, collaborative learning, and individualized instruction (Sergis, Zervas, & Sampson, 2014). Incorporated into a teacher education program, our instructional designs aim to offer a realistic pedagogy and relevant educational knowledge to student teachers so they can be prepared for the implement of a similar instructional approach in their own future classrooms.

Recommendations for instructional designs in flipped classrooms

Currently, empirical research offers no recommendations regarding what technology tools or design features can best support flipped classrooms (Peña & Rosson, 2014). Based on our study results, however, it is indicated that
the student teachers’ learning achievements could be greatly improved when integrating online synchronous group discussion in flipped classrooms. In addition, the integration of the SOP² Model enabled the participants to reach the “Integration” phase, which is a higher level of cognitive presence (Garrison et al., 2001), constituting the second largest proportion in our study. Thus, the SOP² Model, incorporating our instructional designs and procedures, might be able to provide recommendations for future implementations of flipped classrooms in a higher education context.

**Promoting integration and resolution phases for the cognitive presence**

Our study results indicate that the groups with a higher proportion of “Integration” and “Resolution” phases had better learning achievements and progression in terms of their worksheet scores. In addition, both the high- and low-achievement groups had significant correlations between the “Integration” and “Resolution” phases. Thus, it is suggested that more emphasis should be put on facilitating the “Integration” and “Resolution” stages, which are critical factors for increasing participants’ learning achievements. We also discovered that sequential discourse from “Resolution” to “Integration” led to better learning performance than “Integration” to “Resolution.” Thus, to foster “Integration” and “Resolution” phases, several adjustments could be made based on learning activities and instructional design.

Since previous studies have suggested that role assignment could facilitate higher-level cognitive presence and knowledge construction, it is recommended that each group distribute roles, such as group leader, information provider, resolution organizer, and discussion moderator, before online discussion. Meanwhile, instructors could encourage students to conduct small group communication by providing discussion resources such as links, printed flipped materials, PDF files, and personal comments or reflections for each educational context prior to in-class sessions. Participants could then briefly review and become familiar with the prepared materials in order to offer related ideas and thus conduct efficient group discussions. Also, instructors could demonstrate discussion techniques regarding how to reach the “Integration” and “Resolution” phases. For example, each group can re-organize or summarize developed resolutions at the end of their online discussion. In addition, specific cases relevant to the discussed contexts could be offered to show students how to integrate and apply their proposed ideas and resolutions. Lastly, in terms of instructional design, future flipped courses could prolong the duration of synchronous online discussions by offering more time for each discussed context or by reducing the number of discussed questions so it might be more possible for participants to reach the “Integration” and “Resolution” phases.

**Limitations**

The sample size was small, involving a class of 31 student teachers whose learning outcomes might not be generalized to other participants or groups. Since there was no random assignment in this study, the results might be restricted to similar instructional contexts. Even though discourse analysis and evaluation of worksheets might cause subjective bias based on actual situations, we have limited such possibility by having two researchers conduct analysis in order to increase consistency. Although there was no control group, the purpose of the present study was to develop a practical flipped classroom model and thus the focus was on the instructional procedures. In addition, three cycles of the SOP² model were fully conducted throughout the semester, and possible effectiveness was evaluated based on significant improvement in the achievement scores of three different individual worksheets.

**Acknowledgements**

This work was financially supported by the Ministry of Science and Technology, Taiwan, under Grant Nos. NSC 100-2511-S-011-001-MY2, and MOST 103-2511-S-011-006-MY2.

**References**


How to Flip the Classroom – “Productive Failure or Traditional Flipped Classroom” Pedagogical Design?

Yanjie Song1* and Manu Kapur2

1The Education University of Hong Kong, Hong Kong // 2ETH Zurich, Switzerland // ysong@eduhk.hk // manukapur@ethz.ch

*Corresponding author

ABSTRACT

The paper reports a quasi-experimental study comparing the “traditional flipped classroom” pedagogical design with the “productive failure” (Kapur, 2016) pedagogical design in the flipped classroom for a 2-week curricular unit on polynomials in a Hong Kong Secondary school. Different from the flipped classroom where students are provided video clips with new concepts and associated procedures to review at home before solving problems in class, the “productive failure” pedagogical design in the flipped classroom worked the other way around. Supported by mobile technologies, students explored, discussed and solved problems related to the new concepts first in class even though they might come across failures, followed by consolidating the concepts and associated procedures using video clips at home. The pedagogical design is referred to as “productive failure-based flipped classroom” in this study. The study was carried out in two Grade 7 classes: one with “traditional flipped classroom” and one with the “productive failure-based flipped classroom”. Findings show that both classes had significant improvement in procedural knowledge. However, regarding conceptual knowledge, students in the “productive failure” condition performed better than those in traditional flipped classroom. This suggests that the “productive failure-based flipped classroom” pedagogical design may be better able to improve students’ problem solving skills.

Keywords

Productive failure-based flipped classroom, Traditional flipped classroom, Procedural knowledge, Conceptual knowledge, Mathematics learning

Introduction

The flipped classroom has gained prominence in recent years. There is no uniform definition of it. The traditional flipped classroom in this paper refers to the pedagogical design that inverts the teacher’s instruction in the classroom out of formal class time and uses class time for students to actively engage in practice and knowledge construction with technology support (Baepler, Walker, & Driessen, 2014). A review of the literature shows that the approach adopted, in many cases, intends to make use of videos as lecture instruction out of class time, thus affords more classroom time to engage students in active learning (e.g., Charles-Ogan & Williams, 2015; Chen, Yang, & Hsiao, 2015; Jungić, Kaur, Mulholland, & Xin, 2015; Moore, Gillett, & Steele, 2014; Muir & Geiger, 2015; Sohrabi & Iraj, 2016; Wasserman, Quint, Norris, & Carr, 2015). However, it appears that the traditional way of instruction (direct instruction) remains unchanged except that the time spent on the lecturing in class is performed at home. How to enhance students’ conceptual understanding and develop their problem solving skills in the design and implementation of the flipped classroom has rarely been addressed.

How to flip the classroom? The paper reports a quasi-experimental study comparing the “traditional flipped classroom” pedagogical design to the “productive failure” pedagogical design in the flipped classroom for a 2-week curricular unit on polynomials in a Hong Kong Secondary school. Productive failure is defined as “a learning design that affords students opportunities to generate representations and solutions to a novel problem that targets a concept they have not learned yet, followed by consolidation and knowledge assembly where they learn the targeted concept” (Kapur, 2015, p. 52). In the “productive failure” pedagogical design, students explored, discussed, and solved problems related to the new concepts first in class even though they might come across failures, followed by consolidating the concepts and associated procedures using video clips at home supported by mobile devices. The pedagogical design is referred to as “productive failure-based flipped classroom” in this study.

The rest of the paper reviews the literature, followed by research methods. Then the results are presented and discussed.
Relevant literature

“Traditional flipped classroom” pedagogical design

The flipped classroom is also termed as “inverted classroom” or “blended learning” with various definitions (Chen, Wang, Kinshuk, & Chen, 2014). In general, the flipped classroom attempts to free student class time from lectures by providing the new instructional content (including concepts) in the form of video-clips for students to watch as homework; then use class time for active learning where the teacher acts as a facilitator to organize class activities to deepen their conceptual understanding (Roehl, Reddy, & Shannon, 2013). This type of pedagogical design is referred to as “traditional flipped classroom” and is used interchangeably with the flipped classroom in this paper. Active learning is known as any instructional method that engages students in their learning process through collaborative and problem-based learning activities to develop their critical thinking and problem solving skills (Prince, 2004). Although teacher-student contact hours do not change, students can re-play the instructional content at home, and have more teacher-student engagement in class. The ultimate goal is to shift teacher-centered instruction to student-centered learning to change the role of the teacher from a sage on the stage to a guide on the side (King, 1993).

The flipped classroom is gaining popularity. Previous studies have typically been conducted in higher education settings (e.g., Abeyesekera & Dawson, 2015), and in recent years, research on the flipped classroom in school education is on the rise, particularly in mathematics learning and teaching (Muir & Geiger, 2015). The findings are mixed. Some research findings show that flipped learning encourages students to take a more active role in the learning process before and during the class time (Jungić, Kaur, Mulholland, & Xin, 2015); it increases student motivation and satisfaction with their learning (Clark, 2015; Muir & Geiger, 2015; Hernandez-Nanclares & Perez-Rodriguez, 2016; Love, Hodge, Grandgenett, & Swift, 2014); it allows students to learn at their own pace be responsible for their own learning (O’Flaherty & Phillips, 2015; Lai & Hwang, 2016); it provides an opportunity for differentiated teaching for a range of students’ abilities (Herreid & Schiller, 2013); it has the potential to increase students’ learning performance gains (Heyborne & Perrett, 2016); and the affordances of the flipped classroom approach provide an option for mathematics teachers to address the challenges of the twin demands of covering the prescribed curriculum and catering for students’ learning needs (Muir & Geiger, 2015).

However, some studies show that there are no significant improvements of knowledge gains in students’ learning in flipped learning compared to traditional way of learning in mathematics, and students perceptions of the flipped learning are mixed either in school or higher education (Clark, 2015; Love et al., 2014; Wasserman, Quint, Norris, & Carr, 2015). This indicates that flipped classroom may not be applicable to all subjects (O’Flaherty & Phillips, 2015). DeLozier and Rhodes (2016) posit that video instruction as homework itself in the flipped classroom does not appear to result in changes in learning performance, but may provide additional time for in-class active learning activities that enhance learning performance through different learning activities such as student discussions, collaborative activities and presentations. Students particularly appreciate collaboration and instructional video components in the flipped classroom (Love et al., 2014). O’Flaherty and Phillips (2015), reviewing the literature in flipped learning, point out that existing research under-utilizes conceptual frameworks and pedagogical designs in the flipped classroom, and lacks better integration of pre-class activities into face-to-face classes with active learning pedagogies to develop students’ higher order thinking skills such as improved problem solving, inquiry and critical or creative thinking. They also recognize the need for stronger evidence in evaluating student outcomes in the flipped classroom.

Issues

Despite that there are a number of recent studies on the flipped classroom to enhance student active learning, the majority of the studies have reported the practices in higher education and the findings are mixed, and how to design the learning activities to promote active learning is still at a nascent stage (e.g., Abeyesekera & Dawson, 2015; O’Flaherty & Phillips, 2015). Few studies have reported how to develop school students’ problem solving skills and enhance their conceptual understanding in flipped classroom in mathematics inquiry. In addition, the pedagogical design adopted in the flipped classroom, in many cases, intends to be in direct instructional mode, in which students make use of videos as lecture instruction out of class time and practice activities in class time, but are lack of theoretical framework to guide the design and implementation of the learning activities.

In mathematics education, inquiry-based learning is advocated to develop students’ problem-solving skills and enhance their conceptual knowledge (Curriculum Development Council, 2015). Students are typically involved a set of inquiry phases such as engage, explore, explain, extend and reflect processes (Krajcik, Blumenfeld, Marx,
& Soloway, 2000; Marshall & Horton, 2011; Song, 2014). However, in the flipped classroom, the focus is to move lecture tasks to be performed at home, and leave more time for practice and discussion in class. It appears that the traditional way of instruction (direct instruction) remains unchanged except that the time spent on the lecturing in class is performed at home. Thus, existing flipped classroom approach does not seem to leave room for students to “engage” and “explore” the new knowledge/concepts, but transmits the new knowledge to them via video-based micro lectures to be reviewed beyond classroom. This does not mean that direct instruction is useless and does nothing good to learning. On the contrary, “direct instruction does a fairly good job” if the learning goals are oriented to the acquisition of basic knowledge for problem solving without obtaining corresponding conceptual understanding and knowledge transfer (Kapur, 2016, p. 8). However, O’Flaherty and Phillips (2015) postulate that the successful flipped classroom implementation outcomes should take into account effective student learning that fosters problem solving skills and student engagement both within and beyond the class. They call for better integration of face-to-face learning activities and resources with pre-class ones adopting active learning pedagogies in order to motivate students to learn.

In light of the above issues, this study proposes to “restructure” the classroom and explore new ways to flip the classroom in order to enhance their conceptual understanding.

“Productive failure” pedagogical design

“Productive failure” pedagogical design first engages students in unguided problem solving to elicit their prior knowledge, particularly in the failure to solve the problem, followed by using this information to consolidate and aggregate new knowledge (Kapur, 2016). The failure stems from the fact that learners are commonly unable to generate or discover the correct solution to the novel problem by themselves; on the other hand, they are able to generate sub-optimal or even incorrect solutions to the problem, the process can be productive in preparing them to learn better from the subsequent instruction that follows (Kapur, 2014a; Kapur 2014b). This echoes the findings from Granberg’s (2016) study on mathematics problem solving in secondary school education: if students who observe their errors and engage in productive struggles strive for exploring and analyzing the problems in mathematics, their struggles are likely to be productive in constructing new knowledge; on the contrary, if students who observe their errors and engage in unproductive struggles and do not go back to explore and analyze the problems, their struggles may be unproductive.

Kapur (2010) also compared “productive failure” instructional design with a traditional “lecture and practice” instructional design on the curricula unit of rate and speed in a Singaporean secondary school. The research findings show that students from the productive failure condition significantly outperformed their counterparts from the lecture and practice condition in the post test on procedural knowledge, and conceptual understanding and problem solving skills. Similar findings were also obtained in (Kapur, 2014a; Kapur 2014b) study on teaching the topic of concept and procedures of standard deviation, and Kapur and Bielaczyc’s (2012) study on learning the concept of average speed in secondary education.

The aim of this study was to conduct a quasi-experimental study to test the effectiveness of the two pedagogical designs in the flipped classroom, one with “productive failure” pedagogical design, and the other with traditional direct instruction pedagogical design for mathematics learning in a Hong Kong secondary school on developing students’ procedural knowledge and conceptual understanding of mathematical concepts. The next section presents the research methods, followed by the presentation of the research results, and discussions. Finally, a conclusion is drawn from this study.

Research methods

Proposed “productive failure-based flipped classroom” pedagogical design in this study

Premised on inquiry-based learning models (e.g., Hakkarainen, 2003; Krajcik et al., 2000) and adapted from productive failure four core learning design mechanisms (for fuller details, see Kapur & Bielaczyc, 2012), this study proposes a “productive failure” pedagogical design with a two-phase design. They are: Phase I Problem-solving (Engage, Explore and Explain) to be conducted in the classroom in groups; and Phase II Consolidate (video watching) to be conducted out of classroom individually (see Figure 1). In Phase I, “engage” concerns activation and differentiation of prior knowledge related to the targeted concept, “explore” relates to attention to critical conceptual features related to the targeted concept, and “explain” focuses on explanation and elaboration of these features related to the targeted concept. In Phase II, “consolidate” provides students the opportunities for
organizing and assembling features related to the targeted concept. The learning activities are to be carried out in a mobile learning environment where students are encouraged to use their own mobile devices to support their learning both in and out of school and throughout the learning process (e.g., Song, 2014).

The new pedagogy using productive failure pedagogical design is flipped in terms of the learning design where students’ problem-solving goes before the instruction, which contrasts with the learning design of flipped classroom where direct instruction via video clips is assigned as homework task first, followed by more challenging problem solving activities in class. We termed the novel pedagogy using productive failure design in a mobile learning environment as “productive failure-based flipped classroom” which has two connotations: (a) the new design of instruction adopts productive failure strategy; (b) the Consolidate (video-watching) Phase that used to be put at the beginning of the learning activities in flipped classroom using direct instruction is “flipped” to follow the Problem-solving Phase using “productive failure” pedagogical design in the new pedagogy.

![Diagram](image)

Figure 1. “Productive failure-based flipped classroom” pedagogical design

Participants

The study was conducted at a Secondary School with participants of 11 to 12-year olds from two 7th grade classes with the class size of 25 students each. One class adopted “traditional flipped classroom” pedagogical design, which was termed as TFC group; the other class adopted the “productive failure-based flipped classroom” pedagogical design, which was termed as PFFC group. The TFC group had a mean score of 48.35 (SD = 17.38) in the entrance test, while the PFFC group had a mean score of 46.75 (SD = 18.31). The mean scores of the two classes had no significant differences (t = 0.32, df = 48, p > .05). They were students with better learning motivation and higher learning abilities. Each class was divided into 7 groups with 3 to 4 members. Individual differences of the group members were taken into account while grouping the students as some research findings show that mixed ability-based grouping can help improve students’ academic performance (e.g., Boaler, 2008). Two experienced male teachers with around 10 years’ of working experience led the TFC and PFFC groups. Both teachers were enthusiastic about adopting innovative pedagogies in their pedagogical practices and had participated a few seminars and workshops related to flipped classroom pedagogical practices and the productive failure pedagogical design. The concept of polynomials is typically included in the 7th grade curricula unit in mathematics, and therefore, students had no learning experience of the concept.

Learning platform and mobile devices

In addition to the use of customized teaching materials in class, a learning platform – Socrative and Samsung tablets were adopted to engage students in learning. Socrative is a real-time online feedback system where teachers prepared questions beforehand and students took part in in-class activities with their mobile devices. After logging in and answering the questions assigned, students could get immediate results showing the correct answers as well as performance statistics. Twenty-six Samsung tablets were borrowed from school with build-in applications that allow whole-class screen sharing for students to provide immediate response on the monitor for all to view; group screen sharing that enables students from different groups to share their thoughts with one another, and screen monitoring and locking for teachers to keep track of what students were working on.
Research design

The PFFC group adopted the “productive failure-based flipped classroom” pedagogical design in which the first phase involves problem-solving activities in terms of “engage, explore and explain” in class and the second phase took place anywhere outside school where students consolidated the concept concerned with the mathematical concept related to polynomials. In Phase I, exploiting productive failure pedagogical design, the students were encouraged to work in groups trying to come up with different solutions while they were exploring, discussing and sharing in class. The elements of “engage, explore and explain” were used as scaffolds to facilitate collaborative problem solving activities, but without content-related guidance. This was followed by Phase II in which students consolidated the concepts involved in a particular topic through watching video clips at home individually.

The TFC group adopted the “traditional flipped classroom” pedagogical design, reversing the order of direct instruction from “teachers lecturing in class” and “students working on homework at home” to “students watching video clips with instructional content at home” as the first phase and “students practice and solving problems with their team members with teachers’ guidance” as the second phase. The empirical study involved 10 lessons (5 double lessons) spanning over 2 weeks. The pedagogical designs of the two pedagogical learning activities in the TFC group and PFFC group are shown in Figure 2.

![Figure 2. Pedagogical designs for TFC group and PFFC group](image)

Figure 2 shows that the TFC and PFFC groups had the same length of pedagogical practices except for the differences in the time for watching the video clips. The two groups used the same videos consisting of 5 video clips, each with approximately 10 minutes in length. The content of the videos was co-designed by the two school teachers, and recorded using Microsoft Office Mix with various examples to illustrate ideas covered in each lesson. The first video was about the basic concept of algebra, exponential notation, and conversion between text and algebraic; the second went over the basic definition of powers, like terms and unlike terms; the third one was to do with the understanding of monomials and polynomials, arranging in ascending or descending order of powers and addition and subtraction of polynomials; the fourth one is about the multiplication and division of monomials, and the fifth one, also the last one was concerned with multiplication of polynomials. After the students finished watching the video clips, a hardcopy worksheet was distributed to students with a QR code that linked to the video clips on the Socrative learning platform for students’ easy access via QR code reader app.

Before enacting the pedagogical activities, the researcher explained the aims of the study, reviewed the traditional flipped classroom pedagogical design and explained the new pedagogical design with productive
failure via emails to the two teachers. Then we arranged two 2-hour professional development meetings with the teachers on principles of the pedagogical design, and the design of pedagogical activities and methods of carrying out the study.

**Data sources and analysis**

**Data sources and analysis**

Data sources includes pre-, mid- and post-domain tests, the post-conceptual understanding questions, focus group discussions and student survey about their video-clip watching activities regarding location, frequency and devices used. Pre-test was designed to test students’ prior knowledge on the topic of “polynomial” before they started learning the topic with 9 questions; the level of difficulty was low and the content covered the whole unit of polynomial. Mid-test was designed to test students’ mastery of various knowledge related to algebraic after having watched 3 teaching videos (after 6 lessons) with 10 questions; the level of difficulty was fair. Post-test was designed to test students’ understanding of the content knowledge immediately after they learned the whole unit on “polynomial,” especially addition and multiplication of polynomials with 10 questions; the level of difficulty of the test was comparatively higher than the previous two. The pre-, mid and post-tests focused on testing students’ procedural knowledge gains. In addition, in the post-test, 3 conceptual understanding questions were used to test students’ conceptual knowledge gains.

Focus group discussions had three questions: (1) Can you describe the learning activities in learning Polynomials? (2) How do you find the learning activities? Why? (3) What do you think is most important in your mathematics inquiry process? Student survey was conducted in class immediately after the post-test using a hardcopy questionnaire. All students (26 in total) in the TFC group completed the questionnaire; while, in the PFFC group, we collected 21 responses because 4 students provided medical certificates and were absent from class.

**Data analysis**

Both quantitative and qualitative methods were adopted in the data analysis. Table 1 summarizes the data sources and data analysis to address the questions of (a) the effectiveness of the pedagogical design intervention on students’ procedural knowledge gains; and (b) the effectiveness of the pedagogical design intervention on students’ conceptual knowledge gains. SPSS quantitative data analysis software and Excel were adopted in the data analysis. We also chose some examples to demonstrate students’ learning in the two groups. The data related to conceptual understanding questions were analyzed separately from the rest of the post-test questions. Content analysis (Miles, & Huberman, 1994) was adopted for focus group discussions to understand students’ perceptions of the two pedagogical models for their learning.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Data analysis</th>
<th>Effectiveness of pedagogical design intervention on development of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video-clip watching data</td>
<td>Quantitative</td>
<td>(a) procedural knowledge x (b) conceptual knowledge x</td>
</tr>
<tr>
<td>Pre-domain test</td>
<td>Quantitative</td>
<td>x</td>
</tr>
<tr>
<td>Mid-domain test</td>
<td>Quantitative</td>
<td>x</td>
</tr>
<tr>
<td>Post-domain test</td>
<td>Quantitative</td>
<td>x</td>
</tr>
<tr>
<td>Post-test on conceptual understanding</td>
<td>Quantitative</td>
<td>x</td>
</tr>
<tr>
<td>Focus group discussions</td>
<td>Qualitative</td>
<td>x</td>
</tr>
</tbody>
</table>

**Research results**

**Survey results about video-clip watching activities**

In order to compare the learning performance between the TFC and PFFC groups, first of all, we analyzed the survey results about their video clip watching activities, which is shown in Table 2.
Table 2. Results of students’ video watching activities (TFC and PFFC groups)

<table>
<thead>
<tr>
<th>Have you watched the videos?</th>
<th>How many times have you watched?</th>
<th>Where do you watch the videos?</th>
<th>On what device do you watch the videos?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>TFC group</td>
<td>100%</td>
<td>0%</td>
<td>46.2%</td>
</tr>
<tr>
<td>PFFC group</td>
<td>100%</td>
<td>0%</td>
<td>57.1%</td>
</tr>
</tbody>
</table>

Table 2 shows that 100% of the students in both groups watched the video clips, and almost all the students watched them at home. However, regarding “How many times have you watched each video clip?”, it is found that 53.8% of students in the TFC group and 42.9% of students in the PFFC group watched the video clips 2 or 3 times. The χ² test shows that χ² equals to 0.745 (df = 2), p < .05. The results indicate that the two groups have significant difference in the distribution of times in watching the video clips. Another interesting phenomenon is that 27% of students in the TFC group and 47.6% of students in the PFFC group used tablets or desktop computers/laptops to watch videos (χ² = 0.75, p < .05). The difference between the two groups is significant.

Impact of pedagogical design on students’ procedural and conceptual knowledge

To understand the effectiveness of the “productive failure-based flipped classroom” pedagogical design, the results of pre-, mid-, post-test, and conceptual understanding test scores (Full mark is 100 for each) of the TFC and PFFC groups were presented in Table 3.

T-test results show that no significant difference is found between the performance of the two groups on the three time points except conceptual understanding test (Pre-test: t(1,48) = 1.047, p > .3; mid-test: t(1,48) = 1.515, p > .1; post-test: t(1,48) = 0.626, p > .5; conceptual understanding: t(1,48) = -2.089, p < .05). Referring to the mean scores of each group in these tests, there appears to exist an interaction effect between the tests and groups. However, since the contents of the four tests are different, the score difference between those tests may not suggest the same meanings.

Table 3. The mean scores and SD of the four tests

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Mid-test</th>
<th>Post-test</th>
<th>Conceptual understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>TFC group</td>
<td>25.48</td>
<td>14.16</td>
<td>46.74</td>
<td>21.74</td>
</tr>
<tr>
<td>PFFC group</td>
<td>21.84</td>
<td>10.30</td>
<td>37.54</td>
<td>21.63</td>
</tr>
</tbody>
</table>

In order to compare the scores of the four tests in the two groups on a same scale, the original scores were standardized by transforming all of them to T scores, and then used them for further statistical analysis. The T score is calculated based on the Z score, with an overall average of 50, and a standard deviation of 10. Figure 3 shows the mean T scores of the four tests in both groups.

Before transformation, one sample t-test shows that the mid-test scores are significantly higher than pre-test scores (TFC group: F(1,48) = 20.20, p < .01; PFFC group: F(1,48) = 10.74, p < .01); and post-test scores are significantly higher than pre-test scores (TFC group post-test F(1,48) = 5.96, p < .05; PFFC group: F(1,48) = 5.33, p < .05); no significant difference is found in the score of the mid- and post-tests (TFC group: F(1,48) = 1.066, p > .05; PFFC group: F(1,48) = 0.15, p > .05). Given that the level of difficulties from pre-test to post-test is increasing, the results may suggest that both groups show improvement in their learning performance.

After transformation, the data is then analyzed by repeated measures analysis. Mauchly’s test of sphericity (p > .05) shows that the data does not satisfy the requirement of Repeated Measures analysis, and they are applicable for multivariate analysis. The main effect of tests on different time points is not significant as it should be (F(df = 2.8) = 0.00, p > .05), the interaction effect between tests on time points and the pedagogical design is significant (F(df = 2.8) = 6.36, p < .01), and the main effect between the two groups’ pedagogical designs is not significant (F(df = 1) = 0.12, p > .05).

Multivariate analysis of covariance (MANCOVA) test shows that between groups, the T scores of different time points have no significant difference except on the conceptual understanding test (F(df = 1) = 4.33, p < .05). By controlling the pre-test scores, the difference between mid-test, post-test and conceptual understanding tests is
significant \( F(df = 1) = 11.13, p = .01; \) \( F(df = 1) = 6.64, p < .01; \) \( F(df = 1) = 13.68, p < .01. \) Referring to the mean scores of the two groups on each time points, the results indicate that PFFC group performed better than TFC group in the conceptual understanding questions, while TFC group performed better than the PFFC group in the post and mid-tests.

Referring to the mean scores of the two groups on each time points, the results indicate that PFFC group performed better than TFC group in the conceptual understanding questions, while TFC group performed better than the PFFC group in the post and mid-tests.

Since all students in TFC group managed to watch the videos before class and previewed the content to be covered in class (refer to Table 2), this meant that the students had the opportunities to acquire procedural knowledge before in-class learning activities; while students in the PFFC group needed to engage and explore the problems in groups to seek solutions first before consolidating what knowledge they had learned, which might put them at a disadvantage compared to the TFC group. This might explain the TFC group’s relative higher mean score of mid- and post-tests than that of the PFFC group.

However, students in the PFFC group performed better in conceptualizing the like terms and unlike terms. For example, one question concerned with asking student to discuss the difference between like terms and unlike terms. The best conclusion students in TFC group arrived at was, “I think it can be categorized as numbers and letters”; while a student in the PFFC group mentioned, “If the algebraic expressions and their powers are the same, then they are like terms, if not, they are unlike terms.” It is noted that the former conceptualization tended to be too simplistic without giving more concrete ways of distinguishing like terms from unlike ones; while the latter conceptualization was closer to the very definition, i.e., as long as the algebraic expressions and their powers are the same, then they are regarded as like terms.

The results of the three conceptual understanding questions in TFC (with an average correction rate of 52.7%) and PFFC groups (with an average correction rate of 57.6%) are tested by \( \chi^2 \) test. The results show that \( \chi^2 \) equals to 58.3 \( (p < .01) \). This suggests that the learning performance in the two groups is significantly different, and the PFFC group performed better than the TFC group.

**An example case of the development of procedural knowledge in each group**

We chose one student from the TFC group and the PFFC group respectively to show their mathematics knowledge advancement, taking a question from the mid-test and a question from the post-test as examples. The students were asked to “expand and simplify the polynomials.” Figure 4(a) and (b) show the progress of the student in the TFC group from mid-test to post-test. Figure 4(a) indicates that the student understood the multiplication of unlike terms, but did not understand distributive law of multiplication. Thus, he did not understand the theorem that when a term multiplies two terms, they needed to multiply the two terms with the first term respectively, nor did they understand that only addition and subtraction could happen between like terms; and addition, subtraction, multiplication and division could happen between unlike terms. Take the problem of (b) in Figure 4(a) as an example, which shows that ‘3a’ and ‘-8’ are not like terms but the student did subtraction between the two. Figure 4(b) shows that after learning, the student not only understood the distributive law of multiplication, but also grasped the knowledge of the definitions of like terms and unlike terms, polynomial addition, subtraction and multiplication, and positive and negative numbers.
Figure 4. Solutions to expanding and simplifying polynomials in mid- and post-tests by a student in the TFC group

Figures 5 (a) and (b) show the improvement of a student in solving the problems from mid-test to post-test in the PFFC group.

Figure 5(a) demonstrates that the student did not understand the distributive law of multiplication. Take the problem (b) as an example. It shows that the student only did multiplication between ‘10’ and ‘3a,’ but ignored ‘–8;’ the student also worked out the solutions to Polynomial addition and subtraction intuitively. However, Figure 5(b) shows that the student grasped the distributive law of multiplication, in the meantime, understood polynomial four fundamental operations of arithmetic. Moreover, the student had a better and clearer understanding of the concept of the distributive law.

An example case of the development of conceptual knowledge in each group

We choose one conceptual understanding question as an example to show the results in both groups. The conceptual understanding question is as follows:

Given the side length of a square-shaped land is (2a +3) m,
- Determine the perimeter of the square-shaped land in terms of a,
- Determine the area of the square-shaped land in terms of a,
- If there is a triangle-shaped land with area less than that of the square one by (a² + 13a – 7)m², determine the area of the triangle-shaped land in terms of a.

The percentage of correct answers for Parts (a), (b) and (c) are shown in Figure 6 respectively. The $\chi^2$ test shows that overall, PFFC group outperforms TFC group in this conceptual understanding question ($\chi^2(df = 1) = 202.92, p < .01$). The results suggest that the “productive failure-based flipped classroom” pedagogical design is more effective than that of the “flipped classroom” in the improving students’ conceptual understanding and problem solving skills.
Examining further the students’ answers, it was found that 4 students in the TFC group skipped answering the questions in Parts (a) and (b). This was not surprising because in general, Secondary 1 students are usually frustrated with mathematics long/conventional questions; in many cases, without fully comprehending the questions, many of them consider themselves not being able to answer the questions, or even give up (Hong Kong Examinations and Assessment Authority, 2013; 2015). However, in the PFFC group, no student gave up. They invested more effort in seeking the solutions. For example, some students from the PFFC group provided the answer \((2a + 3)^2 m^2\) to question Part (b) using exponential notation which was more concise and appropriate than those who answered \((2a + 3) (2a + 3) m^2\) in the TFC group.

**Relationship between video-watching behaviors and students’ learning performance**

In order to understand whether students’ video-watching behaviors influenced students’ learning performance, one-way ANOVA test was conducted. The results show that there is no significant effect of watching video times with the results of pre-mid-post-tests \((p > .05)\).

**Interview results**

This section reports the results of focus group discussions in both TFC and PFFC groups. Regard Question 1 (Can you describe the learning activities in learning Polynomials?) and Question 2 (How do you find the learning activities?), both groups could clearly describe the pedagogical activities in their classes and expressed their positive attitude toward the pedagogical approaches. As for why they like the approach, students in the TFC group reported that the approach allowed them to consult teachers and ask help from peers in class when they encountered difficulties. They also deemed that the learning activities were interesting which was quite different from the traditional class, where the teacher lectured in class and students did homework at home. Apart from it, they could watch the videos as many times as they needed if they did not understand the learning content. Nevertheless, the most appreciated feature of the approach by the students was that they could do homework in class. One student reported, “In the past, when I came across difficulties in doing my homework at home, I had no one to consult with. Now I can do homework in class and seek help from others if I have problems. In addition, as I need to spend time watching video clips at home, which reduces the time spent on playing online games as I did before.”

While, the students in the PFFC group appreciated the “productive failure-based flipped classroom” because they were provided the opportunities to explore and plan the methods to solve the problem on their own guided by the pedagogical approach, and share their findings in class before the teacher’s instruction. In addition, after their own exploration of the problems, they watched the video clips at home which helped them consolidate what they learned in the inquiry, and left deeper impressions on the misconceptions that they had experienced. A student reported, “I like the pedagogical approach because I have the opportunities to explore and provide solutions first. Even if I make a mistake, I will be able to know why I make the mistake by watching the videos or teacher’s facilitation. Thus, I will not make the same mistake again.”

As for question 3 (What do you think is most important in your mathematics inquiry process?), students in the TFC group considered teacher’s facilitation or peers’ help in the course of problem solving in class was most important. For example a student reflected that “when I did my home work at home, I had no one to consult...
with. Now I can ask my teacher or classmates when I encounter difficulties”; while students in the PFFC group deemed that collaboration was crucial in the complex problem solving process as “a few minds are better than one.”

Discussions

This study shows that both TFC and PFFC groups improved their mathematics knowledge and problem solving skills on the topic of “Polynomial”, and held positive attitude towards their learning experiences. However, it is noted that compared to TFC group, students in PFFC group gained better conceptual understanding of the knowledge related to “Polynomial”; and achieved higher scores in the solving conceptual understanding questions than their counterparts. In addition, their video-watching behaviors and interview results were also different. We discuss these discrepancies in this section.

“Direct instruction” before or after class - Does it matter?

In this study, the TFC group adopted “traditional flipped classroom” pedagogical design, in which students watched video clips to pre-view the new knowledge before they did active learning in class. This means that students received video instruction before they entered into the classroom. While, the PFFC group adopted “productive failure-based flipped classroom” pedagogical design, in which students were engaged in exploring, planning and solving problems first before receiving video instruction for consolidation. Does it matter where to put the video instruction in the pedagogical design?

To delve deep into the pedagogical design, it was found that although both TFC and PFFC groups watched the same video clip, the purposes were different: the “traditional flipped classroom” pedagogical design in this phase did not focus on generating correct solutions in the initial learning stage, but focused on acquisition of basic knowledge of the topic before problem solving (Charles-Ogan & Williams, 2015; Muir & Geiger, 2015); on the contrary, the “productive failure-based flipped classroom” pedagogical design in the “video watching” phase focused on consolidating the concept that students learned in their process of problem solving (Kapur, 2014a; Kapur 2014b). The nature of the learning was divergent: with the former centered on understanding knowledge, and the latter centered on development of conceptual knowledge and knowledge transfer (Kapur, 2016).

The results of the student learning performance show that both groups improved their learning in the mid- and post-tests. However, students in the PFFC group appeared to have better understanding of the concepts than their counterparts. In addition, the results of the conceptual understanding tests show that students in the PFFC group significantly outperformed their counterparts which echoed the research findings in Kapur’s study (2011). They tended to be more confident and willing to face the challenging questions and strive for optimal ways to solve the problems than the TFC group. This indicates that the “productive failure-based flipped classroom” pedagogical design is conducive to cultivate students’ problem solving skills and enhance conceptual understanding. Further, the students in PFFC group watched the video clips significantly fewer times than their counterparts. This might be due to the fact that students in the PFFC group watched the video clips after their inquiry into the problem, and had better or some understanding of the concept, thus made sense of the concept presented in the video clip by just watching it once.

The results of this study imply that it really matters to students’ learning process as for where to put the instructional video clips in the pedagogical design. To understand the nature of students’ prior knowledge is a crucial element in the pedagogical design from a constructivist perspective (diSessa, Hammer, Sherin, & Kolpakowski, 1991). However, in “traditional flipped classroom,” the students’ prior knowledge may not be elicited by watching the instructional video clips solely; while, delaying instructions in “productive failure-based flipped classroom” offers students opportunities to leverage their prior knowledge and the new knowledge under investigation to generate multiple methods for solving a problem (Kapur, 2011).

“In-class learning activities” of the Pedagogical design – Do they matter?

The in-class learning activities designed in the “traditional flipped classroom” and “productive failure-based flipped classroom” were varied. The former one focused on doing exercises and solving students’ own learning problems at their own pace (e.g., Jungić et al., 2015; Lai & Hwang, 2016); the later one laid emphasis on
encouraging students to explore, discuss and work out solutions on their own with the pedagogical guidance (e.g., Westermann & Rummel, 2012), no matter whether their solutions were correct or not (Kapur et al., 2006).

In the in-class activities of the “traditional flipped classroom,” it was noted that students were expected to apply the knowledge acquired from video-watching activity before the class to engaging and discussing more difficult problem solving tasks to cater for their learning needs (Herreid & Schiller, 2013; Muir & Geiger, 2015); while, in the in-class activities of the “productive failure-based flipped classroom,” students were engaged in inquiring into the problems and seeking solutions built on their prior knowledge. Although they might encounter impasse or temporary failures, they were able to overcome these difficulties in the process of exploring and analyzing problems (Granberg, 2016) as well as through teacher facilitation or peer interactions in class. They could also assembly and organize the polynomial concept related features after class through watching the instructional video clips. Kapur (2011) posits that the teacher’s role is critical for students’ learning: students learn when the teacher provides assistance just at the zone of proximal development (Bruner, 1986). Even delaying help may be more productive in developing their epistemic and problem solving skills (Bielaczyc & Kapur, 2010; Kapur, 2011). However, students in the in-class activities in “traditional flipped classroom” might be deprived of the opportunities to activate their prior knowledge by direct video instruction before entering the in-class activities, which were likely to hinder their development of problem solving skills and knowledge transfer in the in-class activities (Kapur, 2016). Thus, the learning activities in the pedagogical design in class do matter to students’ learning.

Limitations

The results of this study cannot be generalized due to a small number of participants involved, and a short study period. In addition, the individual data was aggregated at the group level to address the efficacy of a group as a whole; but the aggregated data might not be able to predict individual learning (Cress, 2008). The results might also be influenced by the two different teachers taking different classes. Moreover, some concerns have not been resolved in this study. For example, in terms of using tablets or desktop computers to watch the video clips, the number of students in the PFFC group was significantly higher than those in the TFC group. Does this difference influence their learning performance? If yes, how? Do we need to prepare different video learning content in “traditional flipped classroom” and “productive failure-based flipped classroom” owning to the different pedagogical designs? How can we better trace and examine the students’ problem solving process and outcomes, and learn from mistakes (Kapur & Bielaczyc, 2012) in “productive failure-based flipped classroom”? These are the questions that deserve to be addressed in future research studies.

Conclusion

This paper reports on a quasi-experimental study of investigating the impact of two different pedagogical designs in flipped classroom on students’ knowledge advancement and development of conceptual understanding and problem solving skills: the TFC group in the “traditional flipped classroom” condition and the PFFC group in the “productive failure” condition. It was found that although both pedagogical designs could help improve students’ procedural knowledge, the PFFC group significantly excelled the TFC group in terms of conceptual understanding and problem solving skills related to “polynomial.” The “productive failure-based flipped classroom” pedagogical design with the transformed strategies of “Productive Failure” and “Flipped classroom” was inspiring for both teachers and students. For teachers, it transforms what used to be a more unidirectional way of teaching in classroom setting into one that places more emphasis on student-centered discovering and inquiring (Krajcik et al., 2000; Marshall, & Horton, 2011). At the same time it “flips” the general idea of having students view instructional videos to learn the content of knowledge at home before class in “flipped classroom” to having students question, think and discover by themselves before watching instructional videos after class so as to consolidate what is learned. This pedagogical approach encourages students to actively question the unknown and find solutions on their own (Kapur, 2014a; Kapur 2014b). It also extends the learning process outside class that allows students to have more time for thinking and discovering in class, thereby promoting their problem solving skills. Thus, the “productive failure-based pedagogical design” indeed addressed the twin demands of covering the mathematics curriculum and motivating students to learn (Muir & Geiger, 2015). Throughout the adoption of this pedagogical design, mobile devices like smart phones played an important, supplementary role to help the students reach their learning goals. Whether the screen size and devices may influence students’ learning performance deserves further research.
References


Empowering Students in the Process of Social Inquiry Learning through Flipping the Classroom

Morris Siu-Yung Jong
Department of Curriculum and Instruction & Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong // mjjong@cuhk.edu.hk

ABSTRACT
The flipped classroom is an educational strategy about inverting the traditional use of in-class time for conducting lower-level learning activities and out-of-class time for conducting higher-level learning activities. Guided social inquiry learning (GSIL), which is a scaffolded constructivist pedagogic approach, has been conventionally adopted in learning and teaching of Liberal Studies (a “young” core senior secondary subject of social and humanities education in Hong Kong). In this research, we aimed to integrate the idea of the flipped classroom into the process of GSIL for promoting students’ learning achievement and self-efficacy in studying Liberal Studies. Apart from delineating the pedagogic design of the proposed “flipped” social inquiry learning (FSIL) approach, this paper also reports our quasi-experimental study on investigating the pedagogic effectiveness of FSIL in comparison with GSIL. There was a total of 215 Grade-11 students from top, middle, and bottom academic-banding schools participating in this study. Results indicated that, compared to the conventional approach, FSIL had different degrees of positive effects on the high, moderate, and low academic-achieving participants. The findings provide grounds for further studying a wider adoption of the flipped classroom in social and humanities education, as well as in constructivist learning and teaching activities.

Keywords
Flipped classroom, Flipped social inquiry learning, Guided social inquiry learning, Social and humanities education

Introduction
The Marzano Learning Sciences Research Lab (Marzano & Toth, 2014) has recently published a research report (namely “Teaching for Rigor: A Call for a Critical Instructional Shift”), revealing that nearly 60% of the classroom time in schools in the United States is still dedicated to direct instruction. The concluding part of the report reminds us again of the importance of learner-centredness in the twenty-first century education.

With the advancement in computing and information technologies, people nowadays tend to automate the things that can be computerised and automatized, and spend more time and effort on those that cannot (Xu, Feng, Zou & Huang, 2012). Regarding conveying information and factual knowledge, there has been evidence that video lectures can be as effective as in-person live lectures (e.g., Cohen, Ebeling & Kulik, 1991; Howland, Jonassen & Marra, 2012; Mayer, 2009; Zhang, Zhou, Briggs & Nunnaker, 2006). Therefore, it has become questionable about the traditional use of face-to-face time inside classrooms for merely “transmitting” and “receiving” subject contents (Bishop & Verleger, 2013). Employing today’s handy video-recording tools and low-cost Internet access, teachers can create online video lectures by themselves and then assign the videos to students as out-of-class “homework.” Hence, the in-class time can be used for conducting higher-order learning and teaching activities that cannot be “automated.” This is the central idea of the flipped classroom — a pedagogic strategy that has recently attracted a lot of educational researchers’ attention.

According to the recent K-12 edition of the New Media Consortium Horizon Report 2015 (Johnson, Adams Becker, Estrada & Freeman, 2015), the flipped classroom is foreseen as one of the most prominent educational strategies in this triennium to transform students from “spoon-fed,” passive learners into self-directed, active learners. In Hong Kong, “exploring the feasibility of introducing the flipped classroom into learning and teaching” is one of the key objectives of the government’s current three-year master plan on information technology in school education (September 2015 – August 2018) (Education Bureau, 2015).

Liberal Studies is a “young” compulsory subject of social and humanities education in the new senior secondary education system under the recent education reform in Hong Kong (Curriculum Development Council, 2014; Education Bureau, 2014). A core objective of Liberal Studies is to develop students’ knowledge of perennial and contemporary issues upon various social and humanities contexts. Guided social inquiry learning (GSIL), which is a scaffolded constructivist learning approach, is conventionally adopted in learning and teaching of the formal curriculum of this subject. The present research aimed to integrate the pedagogic idea of the flipped classroom into the process of GSIL to promote students’ learning achievement and self-efficacy in studying Liberal Studies. A quasi-experimental study was conducted to evaluate the pedagogic effectiveness of this “flipped” social...
inquiry learning (hereinafter referred as FSIL) approach in comparison with the conventional GSIL approach. It involved a total of 215 Grade-11 students from three schools respectively at the top, middle, and bottom academic banding in Hong Kong.

We organise the rest of this paper as follows. The next section is a literature review of the related work. Then, the pedagogic design of FSIL will be elaborated. After that, we will delineate the method, findings, implications, and limitations of the experimental study. At the end of the paper, we will give the concluding observations on this research.

Related work

Guided social inquiry learning (GSIL)

The twenty-first century education encourages students to acquire knowledge constructively on their own rather than receiving it didactically from teachers (Gee, 2013; Prensky, 2012). Inquiry learning is one of the constructivist pedagogic approaches being advocated in the recent decade. It regards learning as the process of seeking knowledge, raising questions, searching for answers, evaluating information, and asking new questions based on new understandings (Elder & Paul, 2005; Wallace & Husid, 2011). Scientific inquiry and social inquiry are two common paradigms of this learning approach. The former has been considerably adopted in science education (de Jong, 2006; Maeots, Pedaste & Sarapuu, 2008). Often, it requires a learner to hypothesise and probe into the “natural” world, look for precise and repeatable evidence for proving his/ her hypothesis, and finally try to draw a unique conclusion that is physically true in the world.

Social inquiry learning, on the other hand, has been notably harnessed in social and humanities education (e.g., Hill, 1994; Jansen, 2011). The inquiry process usually pivots on an open-ended, argumentative societal issue. Instead of looking for the physical truths in the world, it emphasises on a learner’s understanding and reflection on humans and their interactions with the “societal” world from multiple perspectives, values and interests. However, without the provision of necessitated learning support, merely asking school-age students (especially for those who lack prior related experiences) to conduct such inquiry on their own is unrealistic (Jong, 2014; Small, Arone, Stripling & Berger, 2012).

Guided social inquiry learning (GSIL) suggests that there should be appropriate scaffolds to assist students in pursuing the related learning tasks in the course of inquiry (Jong, Lee & Shang, 2013; Milson, 2002). Newell (2009) has proposed a seven-task GSIL model; the tasks include Task Identification, Search Strategy Initiation, Information Location, Information Evaluation, Information Use, Information Communication, and Problem-solving Product Evaluation. Similar task-based elements are also found in other GSIL models such as Big6 (Eisenberg & Berkowitz, 2011), Organized Investigator (Loertscher & Woolls, 2002), etc. However, some researchers (e.g., Chaddock, 2008; Jansen, 2011; Small et al., 2012; Stripling, 2003) have argued that those models are more suitable for supporting the course of information problem solving that engages students in finding the best information about a problem and then re-organizing that information into a product.

Stripling (2008) has suggested that a good GSIL approach should scaffold students in the course of inquiry to explore significant questions upon the societal issue, comprehend the related content and context, develop corresponding arguments on the issue, as well as monitor and reflect on their own learning. Inspired by the notions of Dewey’s (1958) learning through experience, Vygotsky’s (1978) zone of proximal development, as well as Bruner’s (1986) interpretation in learning, she has developed a GSIL model, namely, Stripling Model of Inquiry. This model consists of six guided inquiry phases: Connect, Wonder, Investigate, Construct, Express, and Reflect. Each of these phases is featured with specific actions to support students in carrying out social inquiry learning. Stripling’s GSIL model is one of the popular strategies being adapted and adopted in learning and teaching of Liberal Studies in Hong Kong. The details of this model will be elaborated further in the next sub-section.

Liberal studies

Liberal Studies is a “young” core subject of social and humanities education in the new three-year senior secondary education system under the recent education reform in Hong Kong (Education Bureau, 2014). The implementation of Liberal Studies began in 2009, and the high-stakes public examination (i.e., the Hong Kong Diploma of Secondary Education Examination) started to include Liberal Studies as one of the four core
examination subjects in 2012. Other core subjects include English, Chinese, as well as Mathematics. The statutory curriculum of Liberal Studies underscores the importance of harnessing constructivist strategies to facilitate students in studying this subject, and GSIL is recommended as one of the major pedagogic approaches (Curriculum Development Council, 2014).

The core objective of Liberal Studies is to develop students’ knowledge on perennial and contemporary issues in cultural, social, economic, political and technological contexts (Curriculum Development Council, 2014). The curriculum is composed of three areas: “Self and Personal Development,” “Society and Culture,” and “Science, Technology and the Environment.” Every area consists of a number of thematic modules; each module is organised around a central theme relevant to its area. For example, there are three thematic modules under the Society and Culture area: “Hong Kong Today,” “Modern China,” and “Globalization.” Every thematic module will involve several societal issues. For example, “Disney: Does the dream finally come true in Hong Kong?” is a societal issue involved in the Globalization module. Normally, Liberal Studies teachers use one teaching cycle to cover one societal issue; the duration of a cycle is nine days (see Figure 1). Usually, there are three 70-minute Liberal Studies lessons evenly distributed in each cycle (e.g., on Days 1, 4, and 7). They will assign related homework to students to do during the days that have no Liberal Studies lessons (e.g., on Days 2, 3, 5, 6, 8, and 9).

As aforementioned, Stripling’s (2008) GSIL model is one of the popular pedagogic approaches being adapted and adopted in learning and teaching of Liberal Studies. The implementation of this model in a nine-day teaching cycle is illustrated on the right-hand side in Figure 1. The Connect, Comprehend, and Express phases are implemented during the face-to-face in-class time (i.e., the three Liberal Studies lessons). The pedagogic activities involved in these phases fall into the lower regions of the Bloom’s taxonomy (Anderson et al., 2001), i.e., remembering and/or understanding. The Wonder, Construct, and Reflect phases are implemented during the out-of-class time (i.e., the six days with no Liberal Studies lessons). The pedagogic activities involved in these phases fall into the upper regions of the Bloom’s taxonomy, i.e., applying, analysing, evaluating, and/or creating. The following further elaborate on how a Liberal Studies teacher implements this GSIL approach in a teaching cycle:

![Figure 1. Implementation of GSIL in learning and teaching a societal issue in a teaching cycle](image-url)
Day 1: Connect (the 1st lesson)
The teacher introduces the students to the general background of the societal issue, and equips them with the related key concepts via direct instruction. Usually, the content of the visual aid (the PowerPoint) that the teacher has prepared for the instruction is extracted from the related part of the textbook and/ or the Internet (e.g., images, animations, YouTube videos).

Days 2 and 3: Wonder (out-of-class time)
The students are required to access additional resources related to the issue pre-posted on the school’s LMS (learning management system, e.g., Moodle, Blackboard) by the teacher. Unlike the factual and impartial content encountered in the previous phase, the content of these additional resources, which is extracted from newspapers, anonymous web articles, and Facebook/ Twitter feeds, is more controversial or even irrational in nature. The students are asked to preliminarily analyse and evaluate these “biased” resources, and by the end of this phase, identify some stakeholders of the issue based on the best of their knowledge.

Day 4: Comprehend (the 2nd lesson)
The teacher first asks the students to talk about what stakeholders they have identified in the previous phase. After that, the teacher will debrief the students on why some are regarded as the key stakeholders of the issue (normally 4 to 5) while some are not. Then, the teacher will further elaborate on the basic characteristics and interests of these key stakeholders via direct instruction. And again, the content of the visual aid that the teacher has prepared for the instruction is usually extracted from the textbook and the Internet. By the end of this lesson, each student will be assigned with one key stakeholder role.

Days 5 and 6: Construct (out-of-class time)
Each student is asked to construct his/ her own argument on the issue from the view of the assigned stakeholder role. In order to do this, he/ she should conduct a more in-depth analysis on the issue and apply the knowledge they have gained so far to frame the argument. Also, each of them has to search for information/ grounds (from the textbook and/ or the Internet) to evaluate/ support the rationality and authenticity of the argument, and if necessary, to reshape the argument. Before the end of this phase, everyone needs to create a PowerPoint to document his/ her own constructed argument with grounds.

Day 7: Express (the 3rd lesson)
The teacher invites some students who are playing different stakeholder roles to present their arguments developed in the previous phase. Usually, the teacher will give a debriefing after each presentation. Hence, through the peer sharing, apart from merely standing in the view of his/ her assigned stakeholder role, each student can understand more about different views of the stakeholder roles played by others.

Days 8 and 9: Reflect (out-of-class time)
Each student needs to “blog” a piece of summative reflection (a reflective journal) on the LMS. In the journals, they will reshape their prior arguments developed in the Construct phase (on Days 5 and 6) from a more holistic perspective with a balance of different stakeholders’ views. Also, with the guiding reflective questions pre-posted by the teacher, they will write down what they have learned in the teaching cycle, as well as what, why and how the knowledge gained can be transferred to study other societal issues within the same thematic module and across different thematic modules in Liberal Studies.

Flipped classroom

When discussing the “history” of the flipped classroom, the “definition” given by Lage, Platt and Treglia (2000) is perhaps the most frequently quoted one — “the events that have traditionally taken place inside the classroom now take place outside the classroom and vice-versa” (p. 32). In fact, the early form of the flipped classroom did not necessarily involve videos and the Internet. It was just about a re-ordering of students’ in-class work (listening to the teacher) and out-of-class work (doing homework that assigned by the teacher) in traditional schooling (Baker, 2000). However, there is no denying that using online videos in the flipped classroom is a desirable way to reach today’s video-saturated culture in youngsters’ familiar medium (Bergmann & Sams, 2015).

Bishop and Verleger (2013) have given a more contemporary definition to the flipped classroom — a learning and teaching strategy combining both (i) direct individual instruction outside the classroom via online video lectures, and (ii) student-centred learning activities inside the classroom. This definition generally aligns with
the core pedagogic idea recognised by most educators and researchers who are currently promoting and researching the flipped classroom (e.g., Baeppler, Walker & Driessen, 2014; Bretzmann, 2013; Chen, Wang, Kinshuk & Chen, 2014; Hung, 2015; Lai & Hwang, 2016; Sahin, Cavlazoglu & Zeytuncu, 2015; Wallace, 2014). In fact, both the behaviourist (Skinner, 1968) and constructivist (Piaget, 1970; Vygotsky, 1978) learning paradigms are both found in the flipped classroom. At home, students are required to preliminarily gain foundation knowledge by themselves via accessing online direct instruction videos created by the teacher. These videos engage students in cognitive learning that falls into the lower regions of the Bloom’s taxonomy (Anderson et al., 2001), i.e., remembering and/or understanding. This is the behaviourist part of the flipped classroom. Back to the classroom, students will deepen their knowledge gained at home via pursuing various teacher-facilitated learner-centred tasks. The tasks engage students in cognitive learning that falls into the upper regions of the Bloom’s taxonomy, i.e., applying, analysing, evaluating, and/or creating. This is the constructivist part of the flipped classroom.

In the recent half-decade, on various educational blogs/social media platforms as well as in many media interviews related to education over the world, a considerable number of educators and teachers have shared their observations on students’ learning improvement after the adoption of the flipped classroom (Bretzmann, 2013; Kelly, 2014; Kirch, 2014; Creative Classroom Lab, 2013; Flipped Learning Network, 2014). Nevertheless, there have been critiques that those positive claims are just “experience sharing” rather than vigorous research evidence (Bishop & Verleger, 2013; Handan, McKnight, McKnight & Arfstrom, 2013). In this couple of years, however, empirical studies on the flipped classroom are noticeably on the increase. For example, Baeppler et al. (2014), Sahin et al. (2015), and Zummo and Brown (2016) have examined the adoption of the flipped classroom respectively in learning and teaching of Chemistry, Mathematics and Biology. In those studies, the experimental group students (under the flipped classroom setting) significantly outperformed the control group students (under the traditional teaching setting) in terms of academic achievement and learning motivation. Hung’s (2015) work has been another successful flipped classroom instance in language learning. With a similar experimental-control setting, she conducted a study on harnessing the flipped classroom in learning and teaching of English in which she also obtained similar significant and positive results.

While reviewing the recent flipped classroom research, it is not difficult to notice that most of the studies have pivoted on a very similar focus, i.e., to evaluate whether the pedagogic effectiveness of the proposed flipped approaches is better than the traditional classroom teaching approaches’. There has been very little attention on harnessing the idea of the flipped classroom to improve or enhance existing constructivist approaches to learning and teaching (Jong & Shang, 2016). Also, there is an apparent lack of instances of adopting the flipped classroom in social and humanities education. Teachers in the related subjects often argue that they have used constructivist pedagogic approaches (e.g., GSIL, project-based learning) in practice for years. They do not observe any critical needs to harness the flipped classroom in order to make their teaching more student-centred (Cheung & Jong, 2016).

Some teachers have even suggested that the pedagogic idea of the flipped classroom is incompatible with constructivist education. They deem that “front-loading” the learning process with direct instruction videos will “rob students of the opportunity to explore knowledge on their own” (Bergmann & Sams, 2015, p. 84). Nevertheless, notwithstanding the course that students undergo to reach their learning conclusions by themselves can be beneficial, along the learning process they may encounter fake information, gain inaccurate knowledge, and in turn come to invalid conclusions (Howland et al., 2012; Jong & Tsai, 2016). In fact, flipped videos will not “pollute” the course of constructivist learning; on the contrary, the videos can serve as solid learning scaffolds to support students in their knowledge assimilation and construction process (Cheung & Jong, 2016; Jong & Shang, 2016).

Flipped social inquiry learning (FSIL)

The most vital pedagogic benefit of the flipped classroom is reinventing the in-class time in schools (Hwang, Lai & Wang, 2015). We have proposed to integrate the pedagogic idea of the flipped classroom into the process of conventional GSIL, namely flipped social inquiry learning (FSIL), for promoting students’ learning achievement and self-efficacy in studying Liberal Studies. Specifically, the present study aimed to answer the following two research questions:

- Can FSIL promote students’ knowledge acquisition in comparison with GSIL?
- Can FSIL promote students’ self-efficacy in comparison with GSIL?
Our proposed FSIL approach still employs Stripling Model of Inquiry (Stripling, 2008) as the foundation. Figure 2 shows the contrast between the design and implementation of FSIL (on the right-hand side) and GSIL (on the left-hand side) in learning and teaching of a societal issue in a nine-day teaching cycle (consisting of three 70-min Liberal Studies lessons). FSIL moves the inquiry phases conventionally taking place inside the classroom in GSIL to outside the classroom, and vice-versa.

![A Teaching Cycle](image)

**Figure 2. Contrast between the implementation of FSIL and GSIL in learning and teaching of a societal issue**

As illustrated in Figure 2, different from GSIL, FSIL implements (a) the **Connect, Comprehend, and Express** phases that involve lower-level pedagogic activities (i.e., remembering and/or understanding) outside the classroom (i.e., on the six days with no Liberal Studies lessons), and (b) the **Wonder, Construct, and Reflect** phases that involve higher-level pedagogic activities (i.e., applying, analysing, evaluating, and/or creating) inside the classroom (i.e., during the three Liberal Studies lessons). The following will further elaborate on how a Liberal Studies teacher implements the FSIL approach in a teaching cycle:

**Days 1 and 2: Connect (out-of-class time)**

The students are required to conduct individual learning through watching the teacher-created instructional videos pre-posted on the school’s LMS by the teacher. Usually, the content of the videos is presented in a direct instruction manner, aiming to introduce the students to the general background of the societal issue and equip them with the related key concepts.

**Day 3: Wonder (the 1st lesson)**

The students are divided into groups. The typical class size in Hong Kong schools is from 32 to 38; usually, the teacher will divide the class into 7 to 8 groups (4 to 5 persons per group). Every group is provided with additional resources related to the issue. Unlike the factual and impartial content encountered in the previous phase, the content of these additional resources, which is extracted from newspapers, anonymous web articles, and Facebook/Twitter feeds, is more controversial or even irrational in nature. Each group is asked to primarily analyse and evaluate these “biased” resources with the members, and to identify some stakeholders of the issue based on the best of their knowledge. During the lesson, the teacher walks around the classroom, and if
necessary, will enact just-in-time interventions to assist the groups in conducting the discussion. The interventions can be, for example, providing additional information, clearing up misconceptions, playing the mediator role to resolve the conflict between the members, and ensuring that each member has a chance to elaborate his/her view on the issue. In the second-half of the lesson, the teacher will ask each group to talk about what stakeholders they have identified during the discussion. After that, the teacher will debrief the students on why some are regarded as the key stakeholders of the issue (normally 4 to 5) while some are not. By the end of the lesson, each group will be assigned with one key stakeholder role.

Days 4 and 5: Comprehend (out-of-class time)
There are 4 to 5 different sets of additional resources pre-posted on the LMS by the teacher. Each resource set contains different content which describes the characteristics and interests of a particular key stakeholder. Every student is required to conduct individual learning through accessing the resource set corresponding to the stakeholder role assigned to his/her group. Usually, the resources are presented in a direct instruction manner, in the form of texts (extracted from the content of the textbook), webpage links, and videos (teacher-created and/or YouTube videos).

Day 6: Construct (the 2nd lesson)
Each group is asked to construct an argument on the issue by standing in the view of their assigned stakeholder role. In this phase, the members in each group work together to conduct a more in-depth analysis on the issue and apply the knowledge they have gained so far from the previous phases to frame their group-based argument. Also, each of them has to search for information/grounds (from the textbook and/or the Internet) to evaluate/support the rationality and authenticity of the argument, and if necessary, to reshape the argument. During the lesson, every group is provided with an Internet-connected tablet, and the teacher will intervene the groups whenever necessary (as in the Wonder phase).

Days 7 and 8: Express (out-of-class time)
On Day 7, each group has to (a) record a short video to express the argument with grounds on the issue which has been developed in the previous phase, and (b) post it on the LMS. On Day 8, all students are required to watch the videos which have been posted by other groups. Hence, through this inter-group sharing, apart from merely standing in the view of the assigned stakeholder role, each group can understand more about different views of the stakeholder roles played by others.

Day 9: Reflect (the 3rd lesson)
The students work in groups again to reflect on the limitations of their prior arguments developed in the Construct phase (on Day 6) from a more holistic perspective with a balance of different stakeholders’ views, and reshape the argument. During the discussion, the groups can seek further help from the teacher. In the second-half of the lesson, the teacher will ask each group to present the finally constructed argument and invite other groups to comment on it. Before the end of the lesson, the teacher will give a summative debriefing for the class on what knowledge has been covered in the teaching cycle, as well as what, why and how the knowledge gained can be transferred to study other societal issues within the same thematic module and across different thematic modules in Liberal Studies.

Method
We employed a quantitative approach to answering the research questions. The collection of qualitative data was to serve the purpose of supplementing the quantitative findings.

Participating schools and students
Secondary schools in Hong Kong are categorised into three academic bands based on the overall academic achievement of their students; Band A, Band B, and Band C are respectively the top, middle, and bottom bands. We aimed to select one school at each academic banding from our research partner schools to participate in this study (i.e., three schools in total). The following were the selection criteria:

- The schools have not adopted the flipped classroom strategy in learning and teaching of any subjects.
- The schools are using the same Liberal Studies textbook and the same type of LMS.
- The targeted students (at Grade-11) of the schools have yet to be taught the Globalization module in Liberal Studies.
Finally, three schools respectively at different academic banding were chosen. At each school, we randomly selected two Grade-11 classes. We mixed the two classes, ranked them according to their Liberal Studies examination scores in the previous school-term, and then assigned them to the experimental group and control group alternately. The total number of students involved was 215 (73 from the Band-A school, 72 from the Band-B school, and 70 from the Band-C school). Their average age was 16.74. Table 1 shows the number of students in the experimental and control groups respectively at each school.

<table>
<thead>
<tr>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band-A school</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>36</td>
</tr>
<tr>
<td>Band-B school</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>Band-C school</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

### Experimental procedure and data collection

We conducted the same quasi-experiment at each participating school separately. It involved learning and teaching of a part of the Globalization module (under the Society and Culture area in the Liberal Studies curriculum) via inquiring into a societal issue, “Disney: Does the dream finally come true in Hong Kong?” in a nine-day teaching cycle. The experimental manipulation was FSIL, while the control manipulation was GSIL, as illustrated in Figure 2. Four months before the study, we had started preparing the learning and teaching materials for both the experiment and control manipulations. These materials included the flipped videos and other resources to be pre-posted on the LMS for implementing FSIL (as aforementioned in the section of “Flipped social inquiry learning (FSIL),” as well as the PowerPoints and other resources to be pre-posted on the LMS for implementing GSIL (as aforementioned in the sub-section of “Liberal Studies” under the section of “Related work”). In addition, we set up a review committee to scrutinise all materials developed for both the experimental and control manipulations so as to ensure that the materials were of comparable quality and aligned with the curricular aim of the Globalization module. The committee was composed of two education professors respectively from two other institutions, six Liberal Studies teachers respectively from six non-participating schools (2 at Band A, 2 at Band B, and at 2 at Band C), and one government curriculum officer.

To eliminate the “teacher effect,” we did not request the Liberal Studies teachers from the three participating schools to play the teacher role in this study. Instead, we invited a female Liberal Studies educator, Iris (pseudonym), who was working at another institution to be “the teacher” in the experiment carried out at each school. Iris was well familiar with the Liberal Studies curriculum, the conventional GSIL approach, and the type of LMS (Moodle) that the three participating schools were using. Two months before the study, we had started explaining to Iris the operational details of the research, familiarising her with the pedagogic design and implementation of FSIL, as well as discussing with her the usage of the FSIL and GSIL materials developed for this research.

At each school, we used twelve days in total to complete the experiment and data collection work with the following procedure:

- Iris took nine consecutive days (i.e., a normal teaching cycle for covering a societal issue) to implement the pedagogic part of the experiment, as illustrated in Figure 2. For the experimental manipulation (FSIL), the out-of-class learning and teaching activities were conducted on Days 1, 2, 4, 5, 7, and 8, and the in-class learning and teaching activities were conducted on Days 3, 6, and 9. On the other hand, for the control manipulation (GSIL), the out-of-class learning and teaching activities were conducted on Days 2, 3, 5, 6, 8, and 9, and the in-class learning and teaching activities were conducted on Days 1, 4, and 7.
- Three days after the pedagogic part of the experiment, we administered the knowledge test (1 hour) and self-efficacy survey (10 minutes) to both the experimental and control groups (see the next sub-section for the details of the test paper and questionnaire). Both the test and survey were “unseen” ones, i.e., the students had not been told beforehand that they needed to do so.
- Right after the test and survey, we randomly selected three students from the experimental group for a 30-minute group interview. We aimed to gain more understanding of how they perceived FSIL. The interview was conducted in Chinese (the mother tongue of the students).
Knowledge test and self-efficacy questionnaire

Both the knowledge test paper and self-efficacy questionnaire were in Chinese. The knowledge test was in the typical format of the Liberal Studies public examination. It consisted of three short questions (10 marks each) and one long question (20 marks), i.e., the perfect score was 50. The test questions were customised from the questions related to the Globalization module in past public examination papers from 2012 to 2015. The validity of the test and marking scheme were scrutinised by the review committee (as aforementioned in the previous sub-section). All completed test papers were individually marked by three trained markers who were postgraduate education students majoring in Liberal Studies at our university. The marking was done in an anonymous manner, i.e., the school and group information on the test papers were removed before being passed to the markers. The author of this paper was responsible for discerning any discrepancies in the markers’ work.

The self-efficacy questionnaire (Cronbach’s alpha: 0.92) that we adopted and customized in this study was originally developed by Wang and Lin (2007) and further revised and validated by Wang and Hwang (2012) to measure learners’ belief in their own ability to pursue the required learning tasks for attaining the assigned learning goals. The customised questionnaire contained eight items (e.g., “I realise that I understand the most complex part of this societal issue”). Each item was accompanied by the five-point Likert-scale, ranging from “1” (strongly disagree) to “5” (strongly agree).

Results

For writing convenience, we use School A, School B, and School C to denote the Band-A, Band-B, and Band-C schools respectively. After the three experiments, we received totally 213 sets of completed knowledge test papers and self-efficacy questionnaires; 72 from School A, 72 from School B, and 69 from School C. The return rates were respectively 98.63% (at School A), 100% (at School B), and 98.57% (at School C). (Remark: One student in School A and one student in School C were absent on the dates that we conducted the test and survey). The research findings obtained from each school are presented in the following sub-sections.

Findings at School A

The participants in School A were high academic-achieving students. Table 2 shows the descriptive statistics of the results of the knowledge test and self-efficacy survey obtained respectively from the experimental and control groups. An independent samples t-test on the knowledge test results indicated that the experimental group’s average score (39.45) had no significant difference from the control group’s (38.53), t(70) = 1.17, p > .05. In other words, in terms of knowledge acquisition, the FSIL students did not significantly outperform the GSIL students.

Table 2. Results of the knowledge test and self-efficacy survey in School A

<table>
<thead>
<tr>
<th></th>
<th>Experimental group (N = 36)</th>
<th>Control group (N = 36)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge test</td>
<td>Average 39.45</td>
<td>38.53</td>
</tr>
<tr>
<td></td>
<td>Standard deviation 9.19</td>
<td>8.98</td>
</tr>
<tr>
<td>Self-efficacy survey</td>
<td>Average 4.38</td>
<td>4.35</td>
</tr>
<tr>
<td></td>
<td>Standard deviation 1.11</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Table 3. Males'/females' knowledge test and self-efficacy survey results in School A's experimental group

<table>
<thead>
<tr>
<th></th>
<th>Male (N = 17)</th>
<th>Female (N = 19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge test</td>
<td>Average score 38.89</td>
<td>40.01</td>
</tr>
<tr>
<td></td>
<td>Standard deviation 9.38</td>
<td>9.00</td>
</tr>
<tr>
<td>Self-efficacy survey</td>
<td>Average score 4.36</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>Standard deviation 1.09</td>
<td>1.13</td>
</tr>
</tbody>
</table>

The Cronbach’s alpha value of the self-efficacy survey data was 0.89, indicating the self-efficacy questionnaire maintained its high reliability in the experiment. An independent samples t-test on the survey results showed that the experimental group’s average self-efficacy score (4.38) had no significant difference from the control group’s (4.35), t(70) = 1.15, p > .05. In other words, no significant evidence supported that FSIL could better promote these high academic-achieving students’ self-efficacy in the course of inquiry. Table 3 shows the descriptive statistics of the test and survey results with respect to the males and females in the experimental group in School A. We further conducted two independent samples t-tests to assess whether the gender factor had an influence on
the test and survey results obtained from the experimental group. The analyses indicated that there were no significant differences ($p > .05$).

In summary, according to the quantitative results, the pedagogic effectiveness of FSIL in School A (a top academic-bandung school) was not better (or worse) than GSIL’s. However, in the group interview with the experimental-group students, they shared some advantageous learning experience in the course of FSIL in comparison with the conventional approach to studying Liberal Studies. The following are some translated interview excerpts:

- **Student A1:** At the beginning, the teacher required us to watch a number of videos which covered the general background and basic knowledge of the societal issue (i.e., in the Connect phase, on Days 1 and 2, out-of-class time), and later, to watch a number of videos which covered some specific information about the assigned stakeholder role (i.e., in the Comprehend phase, on Days 4 and 5, out-of-class time) ... It was a good way to help us quickly grasp and remember the key contents. It became our prior knowledge to go further when doing the group work in the classroom ... when watching the videos, I could stop and rewind them whenever needed ..., and I could watch as many times as I needed ... until I fully understood all the things ... Absolutely, I much more prefer watching videos than reading text-based teacher notes.

- **Student A2:** The time for us to construct our initial argument (i.e., in the Construct phase, on Day 6, in-class time) was obviously shorter than the corresponding time used in the conventional inquiry approach (cf., the control group’s Days 5 and 6, out-of-class time). But less was more, because in this teaching cycle more and more student-student and student-teacher interactions took place inside the classroom. The interactions made the whole inquiry process more engaging. I hope my own Liberal Studies teacher can also adopt this approach in his teaching practice.

- **Student A3:** Expressing our group’s argument via video-recording (i.e., in the Express phase, on Day 7, out-of-class time) was innovative. It was much more interesting than creating a boring PowerPoint file for giving a traditional group presentation ... Also, watching other groups’ video-presentations (i.e., in the Express phase, on Day 8, out-of-class time) was an effective way to let us understand more about other stakeholders’ views ... before attending the last lesson, every member already got some substantial ideas about how to revise and finalise the group’s argument (i.e., in the Reflect phase, on Day 9, in-class time), and thus, we could start the discussion immediately when we caught up in the classroom.

**Findings at School B**

The participants in School B were moderate academic-achieving students. Table 4 shows the descriptive statistics of the results of the knowledge test and self-efficacy survey obtained respectively from the experimental and control groups. An independent samples $t$-test on the knowledge test results indicated that the experimental group’s average score (33.78) was significantly different from the control group’s (24.81), $t(70) = 7.12, p < 0.001$. The Cohen’s $d$ was 1.03. In other words, in terms of knowledge acquisition, the FSIL students significantly outperformed the GSIL students. The effect size was large (Cohen, 1998).

| Table 4. Results of the knowledge test and self-efficacy survey in School B |
|-----------------------------|-----------------------------|-----------------------------|
|                             | Experimental group ($N = 36$) | Control group ($N = 36$) |
| Knowledge test              |                             |                            |
| Average                     | 33.78                       | 24.81                      |
| Standard deviation          | 10.05                       | 9.21                       |
| Self-efficacy survey        |                             |                            |
| Average                     | 4.08                        | 3.29                       |
| Standard deviation          | 1.05                        | 0.93                       |

The Cronbach’s alpha value of the self-efficacy survey data was 0.95, indicating the self-efficacy questionnaire maintained its high reliability in the experiment. An independent samples $t$-test on the survey results showed that the experimental group’s average self-efficacy score (4.08) was significantly different from the control group’s (3.29), $t(70) = 3.34, p < 0.001$. The Cohen’s $d$ was 0.82. In other words, there was significant evidence supporting that FSIL could better promote these moderate academic-achieving students’ self-efficacy in the course of inquiry. The effect size was large (Cohen, 1998). Table 5 shows the descriptive statistics of the test and survey results with respect to the males and females in the experimental group in School B. We further conducted two independent samples $t$-tests to assess whether the gender factor had an influence on the test and survey results obtained from the experimental group. The analyses indicated that there were no significant differences ($p > .05$).
In summary, according to the quantitative results, the pedagogic effectiveness of FSIL in School B (a middle academic-banding school) was better than GSIL’s, with a large effect size. In the group interview with the experimental-group students, they manifested their favourable attitude towards FSIL. They deemed that this pedagogic approach made the whole inquiry process more supportive and engaging in comparison with the conventional approach to studying Liberal Studies. The following are some translated interview excerpts:

- **Student B1:** The videos taught us before attending the lessons (i.e., in the Connect and Comprehend phases, on Days 1, 2, 4, and 5, out-of-class time). I found myself more equipped before attending the class. I had more confidence, and thus I was more willing to engage myself in the group activities inside the classroom …… In normal Liberal Studies lessons, sometime the teacher talks too fast, sometime the content is too much to be digested, sometime the content is too difficult … usually, I don’t dare to ask questions in front of the class … I am afraid that I will look stupid … Honestly, I am a slow learner. But this time I could control the teacher’s teaching pace by moving forwards and backwards the play-back head (on the video player), Haha. That’s cool. I like this learning and teaching mode as I can have more individualised learning in accordance with my learning ability.

- **Student B2:** The group discussion conducted during the first and second lessons was very useful (i.e., in the Wonder and Construct phases, on Days 3 and 6, in-class time). It was quite engaging. For example, in the first lesson, we were asked to work in groups to identify the key stakeholders of the issue. With our collective intelligence, my group could quickly identify the key stakeholders … 4 brains are always better than 1 brain, Haha. The discussion process was quite beneficial …… Conventionally, we do this part individually at home (c.f., the control group’s Days 2 and 3, out-of-class time), and I regard this as the most difficult and frustrating part of the whole inquiry process.

- **Student B3:** We are used to conducting the final reflection individually at home (c.f., the control group’s Days 8 and 9, out-of-class time). However, this time we conducted the reflection together with the teacher and classmates during the final lesson (i.e., the Reflect phase, on Day 9, in-class time). I like this arrangement because the interactive and supportive atmosphere and rapport stimulated me to reflect deeper on what I learned in the course of inquiry. This also helped me better retain the knowledge in my mind. I think the knowledge gained will be useful for inquiring into other societal issues in the future.

Findings at School C

The participants in School C were low academic-achieving students. Table 6 shows the descriptive statistics of the results of the knowledge test and self-efficacy survey obtained respectively from the experimental and control groups. An independent samples t-test on the knowledge test results indicated that the experimental group’s average score (20.35) was significantly different from the control group’s (15.31), t(67) = 4.68, p < 0.001. The Cohen’s d was 0.59. In other words, regarding knowledge acquisition, the FSIL students significantly outperformed the GSIL students. The effect size was medium (Cohen, 1998).

| Table 5. Male’s/ female’s knowledge test and self-efficacy survey results in School B’s experimental group |
|--------------------------------------------------|------------------|------------------|
| Knowledge test | Average | Standard deviation | Male (N = 18) | Female (N = 18) |
| Knowledge test | 33.36 | 10.19 | 34.20 | 9.91 |
| Self-efficacy survey | Average | 4.10 | 4.06 |
| Self-efficacy survey | Standard deviation | 1.10 | 1.00 |

| Table 6. Results of the knowledge test and self-efficacy survey in School C |
|--------------------------------------------------|------------------|------------------|
| Knowledge test | Average | Standard deviation | Experimental group (N = 35) | Control group (N = 34) |
| Knowledge test | 20.35 | 7.89 | 15.31 | 8.04 |
| Self-efficacy survey | Average | 3.08 | 2.54 |
| Self-efficacy survey | Standard deviation | 1.08 | 0.98 |

| Table 7. Males’/ females’ knowledge test and self-efficacy survey results in School C’s experimental group |
|--------------------------------------------------|------------------|------------------|
| Knowledge test | Average | Standard deviation | Male (N = 18) | Female (N = 16) |
| Knowledge test | 19.62 | 7.75 | 21.08 | 8.03 |
| Self-efficacy survey | Average | 3.05 | 3.11 |
| Self-efficacy survey | Standard deviation | 1.12 | 1.04 |
The Cronbach’s alpha value of the self-efficacy survey data was 0.90, indicating the self-efficacy questionnaire maintained its high reliability in the experiment. An independent samples t-test on the survey results showed that the experimental group’s average self-efficacy score (3.08) was significantly different from the control group’s (2.54), t(67) = 5.36, p < 0.001. The Cohen’s d was 0.52. In other words, there was significant evidence supporting that FSIL could better promote these low academic-achieving students’ self-efficacy in the course of inquiry. The effect size was medium (Cohen, 1998). Table 7 shows the descriptive statistics of the test and survey results with respect to the males and females in the experimental group in School C. We further conducted two independent samples t-tests to assess whether the gender factor had an influence on the test and survey results obtained from the experimental group. The analyses indicated that there were no significant differences (p > .05).

In summary, according to the quantitative results, the pedagogic effectiveness of FSIL in School C (a bottom academic-banding school) was better than GSIL’s, with a medium effect size. In the group interview with the experimental-group students, they expressed that they were more motivated to learn in the course of FSIL in comparison with the conventional approach to studying Liberal Studies. The following are some translated interview excerpts:

- **Student C1:** Whenever I look at the Liberal Studies textbook at home, I will fall asleep easily ... When I am in the class, I always look forwards to the recess or lunch time ... however, I can sit in front of my computer or hold my phone for hours for learning new things from YouTube. You know? This is the YouTube generation ... all teaching materials should be presented in video format and put them online so that I can watch them when I have a good study mood. Like this time, I watched all the videos posted by the teacher (i.e., the Connect and Comprehend phases, on Day 1, 2, 4, and 5, out-of-class time). Guess where and when I watched them? Haha, I did it when I was taking the bus to the school in the early morning, or I was lying on my bed at 11 pm before I slept.

- **Student C2:** Although every member of my group did view all the videos pre-posted on the LMS, not everyone can fully understand the content. The teacher joined our discussion during the first and second lessons (i.e., the Wonder and Construct phases, on Days 3 and 6, in-class time), corrected some misconceptions in our minds, and further motivated us via providing us with an alternative way to explore the issue. The interventions were very beneficial ...... this kind of quality teacher-student interactions is very rare in normal Liberal Studies lessons ...... The student-student interactions in my group were beneficial too ....... Conventionally, when I read the more controversial content on the LMS at home (c.f., the control group’s Days 2 and 3), I usually have no clue about how to do the interpretation. Fortunately, this time I worked with my group members (i.e., the Wonder phase, on Day 3, in-class time). We seriously discussed the news and information together before determining the stakeholders of the societal issue ... I did learn a lot from my classmates in this teaching cycle.

- **Student C3:** It was a very interesting learning experience. We created and posted our own videos on the LMS ... and watched other groups’ videos (the Express phase, on Days 7 and 8, out-of-class time). In fact, the time for recording our group video should be more than the time for preparing a PowerPoint file. But we quite enjoyed doing it ...... When we were doing the video, we did rehearse for three or four times. The rehearsal made us remember and understanding deeper the content. Now, I can still recall all the content that we recorded in the video ... Interesting! I could remember all the things spontaneously, not by rote ... I will be very happy if my Liberal Studies teacher from now on can use this approach to teaching all other modules in the syllabus.

**Discussion and implications**

The quantitative data collected via the knowledge tests and self-efficacy surveys in this study showed that, in comparison with GSIL, FSIL had desirable pedagogic effects on the participants from the middle and bottom academic-banding schools (i.e., Schools B and C), but not the participants from the top academic-banding school (i.e., School A). However, the qualitative data gathered via the interviews revealed that FSIL provided the flipped students with more advantageous learning experience both inside and outside the classroom than GSIL did.

Students from top academic-banding schools are used to having stronger cognitive ability and higher self-efficacy in learning (Biggs & Moore, 1993). The ceiling effect refers to the level at which an independent variable (the variable being manipulated) is no longer affecting the dependent variable (the variable being gauged) (Cramer & Howitt, 2004). The participants in School A were high academic-achieving students, and thus the ceiling effect might affect the statistical results of the knowledge test and self-efficacy survey.
Nevertheless, at this stage, we will not simply draw a quick conclusion that FSIL is not suitable for or not worthy of being adopted in top academic-banding schools. Instead, upheld by the positive feedback on FSIL gathered from School A, we will further study whether this approach can better empower “top” students when they are inquiring into some intrinsically complex societal issues identified in the meta-studies on the Liberal Studies curriculum (Fung, Tang & Chan, 2011; Ip, 2010; Ip & Fok, 2010), i.e., to decrease the ceiling effect on the research setting via increasing the cognitive load (Sweller, Ayres & Kalyuga, 2011) in the inquiry process.

In Hong Kong, there are only eight government-funded universities. Every year, only around 18% of Secondary-6 (K-12) students can enter these universities. Not surprisingly, the “winners” are usually the students from Band-A schools, because they are used to largely outperforming the students from Band-B and Band-C schools in the high-stakes public examination, i.e., the Hong Kong Diploma of Secondary Education Examination (Jong, 2015). Indeed, the experimental results obtained from the control groups respectively at Schools A, B, and C echo this phenomenon again (see the right-most columns of Tables 2, 4, and 6). Often, Band-C students and even Band-B students cannot enter the universities to receive tertiary education. In consequence, it hinders their future career development and affects their whole lives. Nevertheless, this research proved that FSIL could significantly empower both the low and moderate academic-achieving participants to attain better learning achievement and gain higher self-efficacy in studying Liberal Studies (a core subject in the public examination), respectively with a medium effect size and a large effect size. Therefore, it is worth additional effort on studying a wider adoption of FSIL in this subject as an “equaliser” for improving the “bottom” and “middle” students’ academic performance.

In general, the positive results of the present research align with the ones obtained in most other recent flipped classroom studies (e.g., Baelper et al., 2014; Hung, 2015; Sahin et al., 2015; Zummo & Brown, 2016). However, it is important to underscore that the insights obtained from this study are different from those studies’. As aforementioned in the literature review, most of those researchers aimed to compare their proposed flipped approaches with the traditional, didactic teaching approaches, while the focus of our work was to integrate the flipped classroom idea into a constructivist pedagogic approach (GSIL) and hence to evaluate whether this flipped approach could yield better pedagogic effectiveness than the original’s. Thus, the findings presented in this paper can provide preliminary grounds for researchers and educators who are interested in further investigating the design, implementation, and evaluation of applying the flipped classroom strategy to enhance other constructivist approaches to learning and teaching (e.g., project-based learning, problem-based learning, game-based learning, experiential learning). Moreover, as discussed in the literature review, many of current successful flipped classroom instances are related to STEM (Science, Technology, Engineering, and Mathematics) subjects. There is an apparent lack of instances of adopting the flipped classroom in social and humanities education (Jong & Shang, 2016). We hope our work can shed light on opening up more research opportunities for harnessing this pedagogic idea in learning and teaching of social and humanities subjects.

Limitations

The sample size of this research (involving only 3 schools and 215 students) had yet to be sufficiently large enough to draw a definite conclusion about the pedagogic effectiveness of the proposed FSIL approach. A piece of further research is to scale up the present study with a larger sample size, involving more students from more schools at different academic banding. In addition, the “teacher” in the three experiments, Iris, was a Liberal Studies educator working in the education department in a tertiary institution. Compared to ordinary Liberal Studies teachers in schools, she was more willing to “take risk” and more eager to try innovative approaches to learning and teaching of this subject. This might also be the reason why she was willing to accept our invitation to participate in the research. If we redo this study but replacing Iris by a randomly selected Liberal Studies teacher, the results discussed in this paper may not be simply replicated.

Another limitation, which is regarded as an unavoidable one, is the Hawthorne effect (McBride, 2013). This effect refers to a phenomenon whereby people improve an aspect of their behaviour in response to the fact of change in their environment, rather than in response to the nature of the change itself. In this study, the novelty of adopting a new learning approach, FSIL, in studying Liberal Studies might lead to the experimental group students’ temporarily increased learning achievement and self-efficacy. Can the positive outcomes sustain? We are interested in further studying the sustainability/ change of the pedagogic effects of FSIL on these flipped students when they learn with the same approach for the second time and third time.
Conclusion

The flipped classroom has been widely regarded as a niche for facilitating a new educational paradigm shift to take place in the current decade (Bretzmann, 2013; Chen et al., 2014; Johnson et al., 2015; Lai & Hwang, 2016; Sahin et al., 2015; Wallace, 2014; Zummo & Brown, 2016). This pedagogic strategy is not simply about “eLearning,” or “learning with online videos at home.” It is about revitalising the classroom time with teacher-supported higher-order constructivist learning activities, and moving the traditional lower-order behaviourist learning activities outside the classroom (Jong & Shang, 2016). Teachers are not replaced by computers, mobile devices or videos in the course of flipped learning; rather, they usher students into a blended learning environment where students take ownership of their learning.

In this study, we integrated the flipped classroom into the process of the conventional GSIL approach to promoting students’ learning achievement and self-efficacy in studying Liberal Studies. The statistical results of the quasi-experimental study indicated that, in comparison with GSIL, FSIL had significantly positive pedagogic effects on the moderate and low academic-achieving participants, but not the high academic-achieving participants. However, the interviews with the flipped participants at different academic banding unfolded that FSIL did provide them with more advantageous learning experience both inside and outside the classroom. The findings provide grounds for a wider adoption of FSIL in social and humanities education, and shed light on applying the idea of the flipped classroom to enhance other constructivist approaches to learning and teaching.

In light of the present research limitations, as aforementioned, we plan to scale up this study with a larger sample size, involving more students from more schools at different academic banding. Also, another piece of further research is to investigate the substantiality/ change of the pedagogic effects of FSIL on the present flipped participants (i.e., the three experimental groups) when they learn with FSIL again in studying other thematic modules in the Liberal Studies curriculum.

Acknowledgements

This research was supported by an internal seed project grant, Research Development Fund, from the Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong. I very much thank the participating schools, students, and teacher (Iris). Also, I very much appreciate the scholarly, pedagogic, and technical advice provided by the members of the ad hoc review committee formed for this study.

References


Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. Eugene, OR: International Society for Technology in Education.


Flipping Business Education: Transformative Use of Team-Based Learning in Human Resource Management Classrooms

Chung-Kai Huang1 and Chun-Yu Lin2*

1Department of International Business, National Taipei University of Business, Taiwan // 2Department of Business Administration, National Taipei University, Taiwan // hck2005@ntub.edu.tw // chunyu@mail.ntpu.edu.tw

*Corresponding author

ABSTRACT

With the globalization of macro-economic environments, it is important to think about how to use instructional design and web-based digital technologies to enhance students’ self-paced learning, stir up learning motivation and enjoyment, build up knowledge-sharing channels, and enhance individual learning. This study experimented with the flipped learning course modules which are grounded in the theory of team-based learning (TBL), in which students participated in the preview activities before class, performed in-class activities, and took part in after-class online discussion via modularized assignments. Two Human Resource Management classes, consisting of 104 students and two professors from two Taiwanese universities, participated in this study. Mixed research methods including surveys and interviews, were used to collect both quantitative and qualitative data. The results show the positive relationships among the students’ perceived team members’ valuable contributions, motivation, enjoyment, and learning outcomes. The research findings and implications will provide business education faculty, researchers and decision-makers with ideas and insights on the application of flipped-classroom team-based learning modules.

Introduction

With the developing phenomenon of globalization, advances in information technology, and managerial innovation, the boundaries, scope, and concerns of business education have broadened its horizons (Escobar-Rodriguez & Monge-Lozano, 2012). As they are undergoing a globalization process, higher education institutes worldwide are seeking new business curriculum design which is aligned with the needs of contemporary teaching and learning as well as with the competencies and protocols of workplaces (Avis, Fisher, & Thompson, 2014; Conrad & Dunek, 2012). As the curriculum is highly influenced by its social, economic, and cultural environment, the abilities to look for necessary information, share knowledge, and work collaboratively with others for problem solving in teamwork are particularly desired. Thus, universities nowadays are expected to, and have an urgent need to provide transitional education that highlight these issues and help students face today’s complex changing world (Pellegrino & Hilton, 2013).

In today’s digital age, it is clear that not only is the creation of curriculum innovation and structures important, but it is also evident that so is the way we deliver course content. The transformation of curriculum design makes a difference regarding learners’ acquisition of relevant knowledge, competence, skills and other characteristics (Beetham & Sharpe, 2013) for applying to their professional work domains. Business schools that have incentive schemes using the new pedagogical concept known as a flipped approach and web technology intensively have become more innovation-oriented in order to exhibit much higher incidences of structured training across various coursework (Butt, 2014; Findlay-Thompson & Mombourquette, 2014; O'Flaherty & Phillips, 2015). The purposes of flipping the younger generation’s learning attitudes and habits, as well as linking education closer to the world of work, is of great significance in educational practice (Bergmann & Sams, 2014; Berrett, 2012). In this study, we are concerned about business education in Human Resource Management (HRM) because quality courses provide opportunities for students to conceptualize relevant topics in the evolving functions of HRM within today’s organizations and business environment.

To achieve this goal, we propose an instructional design grounded in team-based learning in a flipped learning scenario, which provides students with valuable opportunities for quality learning. The expected benefits to learners who participated in this approach include promoting student-centered learning and collaboration, as well as improving their engagement in their own education and learning achievement. This article first assesses the advantages and limitations of the existing flipped classroom and the rationale for incorporating of team-based learning in business training and development from the perspective of HRM curriculum design, development and assessment. We intend to flip students’ learning in HRM, with the hope of constructing their learning experience in a way that will be conducive for workplaces and also for society as a whole. Experimenting with innovative business curriculum in HRM helps facilitate integrated growth of business students through meaningful and
interactive activities in a flipped learning fashion. This study aims to provide visibility of curriculum design, implementation and assessment and to enable more evidence-based educational research toward a better understanding of how students learn and how they engage in the course.

**Flipped classrooms overview**

**Pedagogical practice**

Traditionally, in order to facilitate large-class communication, most courses are transmitted in a lecture approach so as to deliver relevant course content that allows the instructor to precisely determine the aims, content, organization, pace and direction of a presentation (Struyven, Dochy, & Janssens, 2010). In planning to transform the pedagogical paradigm and practice, the impact of training of university teachers on their teaching skills and their approach to teaching, as well as to the learning guidance of their students cannot be overlooked and thereby should be taken into account (Postareff, Lindblom-Ylänne, & Nevgi, 2007). In essence, flipping a classroom is a form of blended learning through which students learn content online by watching video lectures, usually outside the classrooms, and homework is finished in class with teachers and students discussing and solving questions (Strauss, 2012). Teacher interaction with students is more personalized, and interactive guidance is used instead of one-way directional lecturing (Bergmann & Sams, 2012).

In flipped classrooms, course content is presented in a lecture format with the instructor conveying factual or procedural type of knowledge and information to students in a face-to-face instructional fashion. Presumably, in this learning model, students take a passive role during the face-to-face session (Xie, Debacker, & Ferguson, 2006). In terms of homework assignments, students are expected to apply the information they receive in class to similar or higher cognitive level problem-solving situations or tasks of practice. Classroom instructors have little time to facilitate and help each individual learner and meanwhile, students have no one to turn to for future assistance after they leave the classroom. While listening to lectures and taking tests in class, the traditional pattern of teaching has been implemented in a way to assign students to read textbooks and work on problem sets outside school. Nonetheless, to help students master the learning materials at their readiness level, flipped learning does what it conceptualizes through utilizing technology to deliver subject content knowledge in multiple formats as students work through the designed activities (Fulton, 2012b).

In flipped learning, teachers intend to transform lecture content to homework, which can literally free up valuable class time to generate more active learning and higher-level cognitive tasks (Kennedy, 2012; Roehl, Reddy, & Shannon, 2013). Most teachers who use the flipped approach create video of approximately 10-15 minutes in duration; the format may consist of images, video, narration and interactive manipulatives (Smith & McDonald, 2013; Tucker, 2012). Students first study the topic by themselves, typically using video lessons prepared by the teacher or third parties. That is, students view recorded tutorials or lectures at home, freeing up limited class time for more active learning work and strategies. Online sharing and collaboration among flipped teaching learners are breaking barriers to help students learn the content, increase self-efficacy and construct learning experience through the materials created (Enfield, 2013). This revolutionary instructional design granted educators, who were thrilled about the possibility of creating learning environments conducive to promote individualized and personalized learning, an opportunity to adopt student learning needs as the driver of instruction (Fulton, 2012a). Regardless of what one names it, flipped teaching, learning or inverted classrooms, the major instructional principles behind this form of blended instruction are drawing more attention from education teachers and leaders around the world. By reversing the sequence of instruction, innovative educators are making a change in the way students learn to in order to ensure better use of class time for active and mastery learning.

**Learning environment and dynamics**

Flipped teaching shifts the focus of learning from a teacher-centered to student-centered paradigm, which affirms the effect of self efficacy and self-regulation (Kuiper, Carver, Posner, & Everson, 2015; McLaughlin et al., 2014). A great amount of studies have indicated that active learning promotes student engagement, satisfaction and achievement (Armbruster, Patel, Johnson, & Weiss, 2009; Tune, Sturek, & Basile, 2013). A flipped classroom is an instructional model that makes master learning manageable and sustainable for a large group of students (Bergmann & Sams, 2012) because students work through a self-facilitated trajectory of curriculum design and improvement in predetermined targeted knowledge construction, competence and work skills. In terms of the pedagogical adoption of flipped classrooms, relevant studies have showed various applications of
the flipping approach (Bull, Ferster, & Kjellstrom, 2012; Herreid & Schiller, 2013; Zhang, Wang, & Zhang, 2012). A simple flip is the most basic level—it is characterized by using online videos, digital textbooks or other forms of media to move lecture materials to a home-based learning environment (Nielsen, 2012). After chunking and separating what might have been a long lecture into a much shorter video, students will be asked to cover the teacher-provided materials on their own and kept track of questions and concerns on which they need clarification.

Face-to-face class time can be used to complete follow-up assignments as students come to class, having the pre-class materials prepared. The purpose of switching up homework assignments and class time is to provide more assistance as students demonstrate a lack of understanding and practice the skills set up from the course objectives. Teachers who aim to maximize face-to-face time can incorporate activities whose learning strategies promote higher-order thinking, social learning and 21st century skills (Galway, Corbett, Takaro, Taïryan, & Frank, 2014). Experienced teachers could take full advantage of the online delivery of lecture materials and implement master-level learning in which a student is accountable for operating both the group and individual tasks (Ash, 2012). Typically, managing a flipped mastery classroom requires teachers to deliver course information, assignment, and assessment through a course management system. The advantages of offering online access to content in a self-paced learning environment are that learners can access the information at their own pace and continually reference recorded material homework. The implementation of flipping changes teachers’ roles from that of “sage on the stage” to one of “guide on the side”, allowing them to work with individuals or groups of students throughout the session (Hughes, 2012). Flipping changes the allocation of teacher time. Instead of imparting the initial lesson in person, the teacher tutors the students when they are stuck via complementary techniques, such as differentiated instruction and project-based learning.

In conclusion, flipped learning involves the lower levels of learning in Bloom's taxonomy (Bloom, Engelhart, Furst, & Hill, & Krathwohl, 1956; Anderson et al, 2001), in which remembering and understanding take place outside the classroom at the student’s own pace, as well as the higher levels of learning in applying, analyzing, evaluating and creating. It accomplishes this by having the instructor serve as a guide in the class (Sarawagi, 2013), taking into the consideration of the cognitive processes that students undergo while working with knowledge. The pre-class preparation helps students clarify their thinking about the course subject, while in-class assignments facilitate rich learning interchange and reciprocal feedback. Students assimilate the knowledge by solving problems and doing application work in class, such as data analysis, discussions, debates, or synthesis activities (Kong, 2014). Students learn by doing and asking questions help both the instructor and the student assess understanding. Flipped classrooms free class time for hands-on work and allow students to gain first-exposure learning prior to class. During the class, students manage to help each other as well, a process-oriented model that benefits both the advanced and less advanced learners.

Team-based learning

Team-based learning (TBL) is a type of instructional strategy that helps organize students and place them in their assigned teams based on their skill sets and backgrounds (Michaelsen & Sweet, 2011; Parmelee & Hudes, 2012; Sisk, 2011). Originally a practice in medical and business domains, TBL was further adopted and extended into classroom practice and use (Gomez, Wu, & Passerini, 2009). TBL is a kind of structured small-group learning that promotes student preparation out of class and application of knowledge, skills, abilities and other characteristics in class. Students are strategically grouped into varied teams of 4-7 students based on the learning purposes, and they work collaboratively throughout the coursework. (Michaelsen, Knight, & Fink, 2002) Students preview their predetermined assignments and prepare the needed materials before each unit or module of the course. Consequently, small and structured student teams work together for the entire semester in TBL environments, utilizing class time to discuss reading assignments and apply concepts for problem solving, either individually or collectively (Michaelsen & Sweet, 2008; Parmelee & Michaelsen, 2010). In business courses in higher education, the incorporation of TBL is critical to develop students’ abilities and skills and help them better understand their business professions, organizational behaviors, and career development, while many tasks and projects are performed through the contributions of the team members via knowledge sharing, information and opinion exchange (Boud, Cohen, & Sampson, 2014).

Regarding the philosophy of course design, TBL is an alternative to lecturing in large class settings. TBL can not only transform traditional content with application and problem solving skills, but it can also reinforce the opportunities for learning how to learn, work, interact, and collaborate in a team for accomplishing the required work (Engeström, 1999; Parmelee, Michaelsen, Cook, & Hudes, 2012). TBL shifts the focus away from reducing the amount of traditional classroom lecturing by the professor to increasing the in-class or out-of-the-class
participation and application by student groups (Gregory & Thorley, 2013). To facilitate such instructional design formats, course materials are divided into units or modules while the students are split into teams. There is little lecturing, with the instructor taking the role of a facilitator of teams that are formed at the start of the semester. Most of the learning experiences occur when working in a team during in-class or out-of-class interactions (Michaelsen, et al., 2002). Many schools have adopted some version of TBL due to several of the benefits listed above, and they have found that this kind of design, either on-site or on-line, is conducive to greater long-term knowledge acquisition and retention, as compared to a traditional passive-oriented lecture curriculum (Hwang, Singh, & Argote, 2015; Lee & Lim, 2012; Zhu, 2012).

Based on the aforementioned principles and guidelines, TBL promotes active learning, participation and engagement from individual and group levels (Conway, Johnson, & Ripley, 2010). Implementing traditional face-to-face TBL approaches, nevertheless, requires several prerequisite preparations in order to overcome the challenges (Gomez, Wu, & Passerini, 2010). Also, the amount of time spent on guiding the rules for students’ grouping and readiness in their interactions as well as the time needed from the instructor to explain and clarify learning content and materials in different in-class activities and inter-team dialogues (Jahng & Bullen, 2012). On the other hand, web-mediated technologies resolve the constraints of time and space and support better classroom management practices, since many activities can be executed synchronously or asynchronously at the student’s own pace in a team learning context (Hew & Cheung, 2012). Taking advantage of extending the classroom activities to online space, students can have unlimited access to materials and opportunities to discuss the assigned topics with their team members, thus leading to better consolidated learning outcomes (Salmon, 2012, 2013).

**Instructional design of flipped-classroom TBL**

HRM is one of the fundamental courses for undergraduate business and management majors. In a regular HRM course, students are assumed to obtain the ability to evaluate and apply theories of managerial and social science disciplines to workplace issues (Coetzee & Sillington, 2014). In addition to the building of basic knowledge, students are expected to acquire relevant strategic HRM functional capabilities used to select, develop, and motivate workers as well as cultivate their own analytical, communication, and decision making skills (Beer, 2015). Nonetheless, traditional lecture- and exam-based HRM courses lack this common practice and integrative design, in which students can take the full responsibilities in developing these targeted learning goals and outcomes. In higher education, there has been a lack of empirical studies on flipped-classroom TBL in college level HRM. Situating team-based learning in flipped classrooms, we intended to create more opportunities for students to practice the knowledge and skills through a structured but flexible framework that provides pre-class, in-class, and post-class learning activities. Students explore teamwork and solve scenario- and case-oriented problems by engaging themselves in a process of negotiation and contributing to the provision of constructive feedback to other team members.

The objectives of the course are to provide students with basic understanding of human resource functions, strategic philosophy and the possible organizational roles as business partners, learning specialists, professional specialists, project managers and corporate social responsibility practitioners (Davis, Naughton, & Rothwell, 2004; Jamali, El Dirani, & Harwood, 2015; Lin & Huang, 2015). The design of establishing fixed teams to function as dynamic learning cohorts allows students to develop their capabilities and competencies needed to fulfill responsibility and surmount the possible challenges and emerging issues in the changing business world (Lin, Huang, & Liang, 2015; Huang, Lin, & Kuo, 2015). Thus, this present study sets out to flip the classroom for HRM learners in the TBL format at two universities in Taiwan to better explore its instructional design, examine pedagogical potentiality and contribute to the growing body of research based on the variables explored. The instructional design of this flipped-classroom TBL (FC-TBL) is centered on the philosophy of the constructivist paradigm that encourages students to take an active from a passive role and contribute to developing their own knowledge and information seeking process (Jonassen & Land, 2012).

Table 1 illustrates the different learning phases of FC-TBL design, pre-class, in-class and post-class modules, which formulate the coherent themes within the HRM course. The pre-class module requires individual preparation including the selected videos, cases and assignments to be completed. The main purpose is to help students review class-related materials and share ideas and notes within the team.
Table 1. FC-TBL learning phases per module

<table>
<thead>
<tr>
<th>Module and activity</th>
<th>Materials and assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-class</strong></td>
<td></td>
</tr>
<tr>
<td>1. Instructor assigns videos or reading materials about the chapters’ topics for students to preview through online forum. Videos remain online for consultation through the course.</td>
<td>• English videos (10 minutes)</td>
</tr>
<tr>
<td>2. Students watch the videos and take notes individually.</td>
<td>• Chinese videos (30 minutes)</td>
</tr>
<tr>
<td>3. Students share their ideas, thoughts, and notes about the topics on team.</td>
<td>• Business cases</td>
</tr>
<tr>
<td>4. Teams summarize the key points of videos or reading materials.</td>
<td>• Write individual job resume</td>
</tr>
<tr>
<td>5. Students and teams are well-prepared before the class meetings.</td>
<td>• Write job description</td>
</tr>
<tr>
<td><strong>In-class</strong></td>
<td></td>
</tr>
<tr>
<td>1. Instructor leads case discussion and group presentations.</td>
<td>• Reading materials</td>
</tr>
<tr>
<td>2. Students share their ideas, thoughts, and notes about the topics on team.</td>
<td>• Textbooks</td>
</tr>
<tr>
<td>3. Teams synthesize and organize ideas and present.</td>
<td>• Movie or video clips</td>
</tr>
<tr>
<td>4. Watch movies and group discussion.</td>
<td>• Posters and markers</td>
</tr>
<tr>
<td>5. Individuals interview and group interview.</td>
<td></td>
</tr>
<tr>
<td>6. Teamwork and role play</td>
<td></td>
</tr>
<tr>
<td>7. Teams interview people on campus</td>
<td></td>
</tr>
<tr>
<td>8. Instructors assign readings and team homework</td>
<td></td>
</tr>
<tr>
<td>9. Instructors facilitate and wrap up.</td>
<td></td>
</tr>
<tr>
<td><strong>Post-class</strong></td>
<td></td>
</tr>
<tr>
<td>1. Teams discuss assigned readings and homework.</td>
<td>• Reading materials</td>
</tr>
<tr>
<td>2. Teams work on written paper.</td>
<td>• Reference books</td>
</tr>
<tr>
<td>3. Individuals have discussion on topics in online forum.</td>
<td></td>
</tr>
<tr>
<td>4. Teams interview human resource managers.</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 1. FC-TBL activities](image-url)
After the pre-class module, students rotate on a predetermined fixed schedule between the in-class module and post-class module. The in-class module consists of teacher-guided activities that require students to apply the concepts of knowledge and information previously gained from the pre-class exercise, as well as to perform a task or solve a problem in a new situation. In the post-class module, students are exposed to different Facebook-supported online discussions or projects, in order to integrate the course chapter materials as end-of-the-unit assessments. That is, students manage to engage in varied pre-class, in-class and on-line activities, such as taking initiatives and challenges in multiple ways for their own assigned tasks and project learning, facilitating discussions among their peer learners and instructors, and stimulating critical thinking in order to achieve higher-level learning.

Figure 1 presents a visualized depiction of the major aspects of FC-TBL activities and the details of the activity flow within a flipped learning environment. While both individual and group activities were introduced, students gained the opportunities to engage in active learning while preparing for their assigned readings, videos, sharing notes and posting responses and feedback for the discussion prompts.

Research methods

Participants

Participants were 120 business-majored undergraduate students at two National universities in Taiwan. All of them took the HRM courses based on the flipped-classroom team-based learning models which were designed by our research team in the same semester. A total of 104 surveys were collected and most of them were valid for data analysis. The response rate was 86.67%. On the other hand, qualitative data were collected in order to supplement the quantitative findings by interviewing six volunteered students and two instructors based on the convenience sampling.

Measures

The previous studies had generated fruitful findings and knowledge about learners’ perceptions about computer-supported team-based learning (CS-TBL). Considering the applicable research purpose and scope, the evaluation framework and measures from Gomez, Wu, & Passerini (2010) were adopted and modified in this study to evaluate learners’ perceptions, feelings, and learning outcomes in the flipped team-based learning setting. As their research was originally conducted in the CS-TBL environment with computer-mediated learning modules, modifications to the survey items were made to specify the FC-TBL by fine-tuning a few context-specific words. By doing so, the authors attempted to ensure that the content validity are maintained as high as possible grounded in the current research design (Adcock & Collier, 2001). In terms of construct validity, the average variance extracted (AVE) was applied to evaluate the discriminant and convergent validity of the measurement (Hair, Black, Babin, Anderson, & Tatham, 2006). Table 2 presents the factor loadings and the AVE values for each construct.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Loadings</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived team member’s valuable contributions (PTMV)</td>
<td>PTMV1</td>
<td>0.894</td>
</tr>
<tr>
<td></td>
<td>PTMV2</td>
<td>0.861</td>
</tr>
<tr>
<td></td>
<td>PTMV3</td>
<td>0.875</td>
</tr>
<tr>
<td></td>
<td>PTMV4</td>
<td>0.910</td>
</tr>
<tr>
<td></td>
<td>PTMV5</td>
<td>0.788</td>
</tr>
<tr>
<td></td>
<td>PTMV6</td>
<td>0.747</td>
</tr>
<tr>
<td>Perceived motivation from FC-TBL (PM)</td>
<td>PM1</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>PM2</td>
<td>0.952</td>
</tr>
<tr>
<td>Perceived enjoyment from FC-TBL (PE)</td>
<td>PE1</td>
<td>0.831</td>
</tr>
<tr>
<td></td>
<td>PE2</td>
<td>0.831</td>
</tr>
<tr>
<td></td>
<td>PE3</td>
<td>0.871</td>
</tr>
<tr>
<td></td>
<td>PE4</td>
<td>0.826</td>
</tr>
<tr>
<td></td>
<td>PE5</td>
<td>0.925</td>
</tr>
<tr>
<td>Perceived learning from FC-TBL</td>
<td>PL1</td>
<td>0.866</td>
</tr>
<tr>
<td></td>
<td>PL2</td>
<td>0.905</td>
</tr>
<tr>
<td></td>
<td>PL3</td>
<td>0.735</td>
</tr>
</tbody>
</table>
In addition, the internal consistency reliability was measured by examining the Cronbach’s alpha. The paper-based questionnaire, consequently, consisted of 18 items and four measures, including perceived team member’s valuable contributions to FC-TBL (PTMV), perceived motivation from FC-TBL (PM), perceived enjoyment from FC-TBL (PE), and perceived learning from FC-TBL (PL). More specifically, PTMV has three items. The sample question is “Most classrooms’ comments are very valuable.” PM has two items. The sample question is “FC-TBL motivated me to do my best work.” PE has five items, and the sample question is “I enjoy sharing my knowledge of course related materials with my team through FC-TBL.” PL has six items. The sample question is “The learning quality of course materials was improved by the team activities.” These first three variables were rated by a 5-point Likert scale, and the last variable was rated by a 7-point Likert scale.

Upon close attention to the literature of flipped classrooms and team-based learning, the research hypotheses that guided this study are formulated as follows:

- **Hypothesis 1a**: Higher perceived team member’s valuable contributions will increase motivation from FC-TBL.
- **Hypothesis 1b**: Higher perceived team member’s valuable contributions will increase enjoyment from FC-TBL.
- **Hypothesis 1c**: Higher perceived team member’s valuable contributions will increase perceived learning from FC-TBL.
- **Hypothesis 2a**: Higher perceived motivation will enhance learning from FC-TBL.
- **Hypothesis 2b**: Higher perceived motivation will enhance perceived enjoyment from FC-TBL.
- **Hypothesis 3**: Higher perceived enjoyment will enhance learning from FC-TBL.
- **Hypothesis 4a**: Perceived motivation mediates the relationship between perceived team member’s valuable contributions and perceived learning from FC-TBL.
- **Hypothesis 4b**: Perceived enjoyment mediates the relationship between perceived team member’s valuable contributions and perceived learning from FC-TBL.

![Figure 2. Flipped-classroom team-based learning (FC-TBL) research framework](image)

**Results**

**Quantitative findings**

Table 3 presents descriptive statistics and inter-correlations of all measured variables. As expected, the associations between all variables are correlated. Perceived team member’s valuable contributions is positively related to perceived learning, motivation, and enjoyment ($r = 0.70, p < .01$; $r = 0.61, p < .01$; $r = 0.67, p < .01$). Perceived motivation is positively related to perceived enjoyment and learning ($r = 0.82, p < .01$; $r = 0.76, p < .01$). Perceived enjoyment is also positively related to perceived learning ($r = 0.77, p < .01$).
Table 3. Descriptive statistics and inter-construct correlations

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>PTMV</th>
<th>PM</th>
<th>PE</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTMV</td>
<td>3.77</td>
<td>.55</td>
<td>.79</td>
<td>.79</td>
<td>.79</td>
<td>.79</td>
</tr>
<tr>
<td>PM</td>
<td>3.57</td>
<td>.65</td>
<td>.61**</td>
<td>.90</td>
<td>.90</td>
<td>.90</td>
</tr>
<tr>
<td>PE</td>
<td>3.55</td>
<td>.57</td>
<td>.67**</td>
<td>.82**</td>
<td>.82**</td>
<td>.82**</td>
</tr>
<tr>
<td>PL</td>
<td>5.05</td>
<td>.88</td>
<td>.70**</td>
<td>.76**</td>
<td>.76**</td>
<td>.76**</td>
</tr>
</tbody>
</table>

Note. *n = 104; *p < .05, **p < .01, ***p < .001; reliability coefficients are in parentheses.

Regression analysis has been used to further examine the relationship among variables in this study. The results show that all hypotheses are supported as shown in Table 4, the summary of hypothesis results. Accordingly, hypotheses 1a, 1b, and 1c were all supported. That is, higher perceived team members’ valuable contributions will increase perceived motivation, enjoyment, and learning from the FC-TBL learning settings. Hypotheses 2a and 2b are supported, which suggests that higher motivation also enhanced learners’ enjoyment and learning in this context. Hypothesis 3 is supported as well, which means that learners’ higher perceived enjoyment enhanced their learning.

Table 4. Summary of hypothesis results

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>β Value</th>
<th>Significant p-value</th>
<th>Supported (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1a</td>
<td>.61</td>
<td>*p &lt; .001</td>
<td>Yes</td>
</tr>
<tr>
<td>H1b</td>
<td>.67</td>
<td>*p &lt; .001</td>
<td>Yes</td>
</tr>
<tr>
<td>H1c</td>
<td>.69</td>
<td>*p &lt; .001</td>
<td>Yes</td>
</tr>
<tr>
<td>H2a</td>
<td>.76</td>
<td>*p &lt; .001</td>
<td>Yes</td>
</tr>
<tr>
<td>H2b</td>
<td>.82</td>
<td>*p &lt; .001</td>
<td>Yes</td>
</tr>
<tr>
<td>H3</td>
<td>.77</td>
<td>*p &lt; .001</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Mediating effects

According to Table 5, Model 4 displays the significant mediating effect of perceived motivation from FC-TBL on the relationship between perceived team member’s valuable contributions and learning from FC-TBL (β = .54, p < .001). In Table 6, Model 8 shows the significant mediating effect of perceived enjoyment from FC-TBL on the relationship between perceived team member’s valuable contributions and learning from FC-TBL (β = .55, p < .001). Thus, Hypothesis 4a and 4b are supported as well. Namely, perceived team member’s valuable contributions can impact perceived motivation and enjoyment, and then, perceived motivation and enjoyment can impact perceived learning from FC-TBL.

Table 5. Mediating effect of motivation

<table>
<thead>
<tr>
<th>Variables</th>
<th>PM</th>
<th>PL</th>
<th>PL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTMV</td>
<td>.61***</td>
<td>.69***</td>
<td>.37***</td>
<td>.37***</td>
</tr>
<tr>
<td>PM</td>
<td>.37</td>
<td>.48</td>
<td>.58</td>
<td>.58</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>.36</td>
<td>.48</td>
<td>.58</td>
<td>.58</td>
</tr>
<tr>
<td>F</td>
<td>59.85***</td>
<td>95.02***</td>
<td>142.08***</td>
<td>100.89***</td>
</tr>
</tbody>
</table>

Note. *n = 104; *p < .05, **p < .01, ***p < .001.

Table 6. Mediating effect of enjoyment

<table>
<thead>
<tr>
<th>Variables</th>
<th>PE</th>
<th>PL</th>
<th>PL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTMV</td>
<td>.67***</td>
<td>.69***</td>
<td>.33***</td>
<td>.33***</td>
</tr>
<tr>
<td>PE</td>
<td>.45</td>
<td>.48</td>
<td>.59</td>
<td>.59</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>.44</td>
<td>.48</td>
<td>.58</td>
<td>.58</td>
</tr>
<tr>
<td>F</td>
<td>82.29***</td>
<td>95.02***</td>
<td>145.59***</td>
<td>92.93***</td>
</tr>
</tbody>
</table>

Note. *n = 104; *p < .05, **p < .01, ***p < .001.
Qualitative findings

In terms of the students’ perceived team member’s valuable contributions, this perception plays a significant role in influencing their perceived motivation. It is believed that the major benefits of working as a team is that team members can take an active interest in supporting each other according to the targeted mission and assigned tasks. Thus, students who take a positive perception of the value of teamwork are more inclined to possess higher motivation towards their learning. As Instructor A reflected,

Throughout this semester, we [the instructor and the teaching assistants] found that the dynamics of a team are infectious. When a student is aware of his/her team members’ expectation of their performance, he/she will adjust his/her own expectations according to that of the group’s in order to live up to the shared standards. The individual contribution leads to a positive reciprocity dynamic within the team, as members tend to perform better to reciprocate the received input with the favorable output.

This idea resonated with Instructor B’s belief in social learning theory.

My original design of flipped learning was based on the idea of encouraging my students to observe and emulate behaviors of others in a team. It’s is similar to the Chinese philosophy that students learn most effectively when they are literally involved in the teamwork and are inspired to contribute to the team learning…It’s a positive and helping cycle.

Yet, Instructor A indicated the negative effects of initiating this flipped project in the HRM course as well.

The flipped learning mode actually requires a workload which compares to the traditional HRM course. In the past, the instructor may only have needed to prepare the lecture materials and take charge of some homework grading. However, in a flipped classroom, more preparation is needed in sorting out relevant supplementary materials, both for the pre-class preview sessions and after-class online discussions. Additionally, the success of the flipped tasks relies heavily on students’ motivation and self-paced learning. If will be difficult if students can’t contribute too much time to the team’s work or lack incentive to fit into this learning model.

Given the benefits of the three phases of course design of a flipped classroom, Instructor B also mentioned that the effort and time he needed to invest in facilitating the online assignments was its drawback. He said,

The workload had drastically increased and thus the support from the teaching assistant became very critical. Sometimes the students were not motivated, I [as the instructor] had to figure out ways to keep encouraging and engaging them…To respond to the online posts was also time-consuming. It took me half a day by providing feedback to a few students’ discussions.

In terms of team-based work, student A expressed the mechanism regarding how his group members cooperated.

In every project, [our team] would distribute the work based on the [team members’] abilities. The leader took the initiatives and the other members assisted each other to solve the problems generated by each of us.

Student A further elaborated on how the valued contribution had an impact on their collective learning motivation. He stated that “…we would ask each other for help, but meanwhile feel embarrassed as well [if the request for help is too much]. Because of our learning characteristics, we’re all independent and cooperative.”

On analyzing the relationship between the perceived motivation and perceived enjoyment, Instructor B shared his postulation, expected outcomes and classroom experience.

In line with the concept of scaffolding, team learning took place when team members finished watching the course videos, further discussed the topics and exchanged ideas with one another. Actually, some classmates had better understanding, whereas others need others’ [students’] help and support. For those who were a bit behind, they could assimilate the main course content through this process…This supporting and participatory curriculum design is good because it not only ramps up the learning motivation and enjoyment but also the enjoyment.

In addition to Instructor B’s analysis, Instructor A took a slightly different perspective by mentioning students’ learning characteristics.
I personally think the perceived motivation would lead to the perceived enjoyment in group work and learning. This is more observable in groups in which students value the quality and outcome of learning...I read students’ online postings on a regular basis and [I] did find students’ cheering comments from time to time...Yet, there were times I discovered students’ low motivation in fulfilling the required work, regardless how devoted and engaged other team members were. It’s disappointing in terms of the class morale and discipline.

Meanwhile, three students indicated their perceived enjoyment through the motivating discussion mechanism out of the FC-TBL. Student A indicated her satisfaction and preference to this type of learning mode. She said, “The [team] production was not limited to scarce output in the textbook. [I] obtained a lot of information.” Student B noted the discrepancy in this collaborative process, but she enjoyed the team’s attempt to “unify the diversity among members into a final consensus.” Adding to the opinions of Students A and B, Student C pointed out the synergy from both individual and TBL. He emphasized that “the support and learning within the group help a lot because of how the consulting partners to brainstorm and discuss together.”

Overall, both teachers and students acknowledged that FC-TBL fosters the learning outcomes by bonding the group members’ interaction as well as spurring up the motivation and engagement internally and externally. As Student D showed in her conclusive evaluation by providing her feedback with the HRM course design:

The course design is very useful. Regarding the preview work, I don’t have any negative feelings, but sometimes I lack the momentum to complete it. The use of [online] platform and assignments improve my learning effectiveness. The adoption of video clips and materials for preview has its advantages. Videos can help me easily capture the chapter’s ideas, but not in-depth. For the English-speaking videos, there are some difficulties in digesting the content, though.

Disliking the traditional teacher-fronted setting, Student A also showed his positive perspectives on the three-phase structured flipped approach. According to his reflection, “I am not good at memorizing, nor do I like lecture-based teaching. Compared to an exam-driven textbook approach, [I] am more into being pushed for absorbing the content little by little.” Student B especially liked the pre-class module because of the required preview duties. He stated that “After watching the video, each group has to set up deadlines, write the summative note and turn in to the group leader for the final check.” He further pointed out that the in-class module was effective because “classmates would proposal questions to each other, seeking the information and finalizing the group consensus via the collaborative discussions.” Learning from the varied perspectives, Student E also highlighted the benefits of in-class group discussion. He emphasized that “by reading textbooks only, we think in a similar way, but by understanding each other’s opinions, we integrate and grow.” As for the post-class module, Student F indicated the capabilities that Facebook provide by saying “…Facebook is useful for information collection and increasing interaction. The group discussion area moderates our work distribution and meetings.” To sum up, the results of the qualitative analysis supplemented the quantitative findings, providing additional insights into the perspectives of instructors and students on FC-TBL.

Discussion

The present study first developed a module of FC-TBL in order to promote active learning and participation of business majors in the courses of HRM at undergraduate level. Subsequently, the well-organized team-based course design, along with pre-class activities, in-class activities, and post-class activities was implemented for two HRM classes for a whole semester. In order to evaluate the learning module, a mixed methods research design was used. We found that learners’ perceived team member’s valuable contributions generally play a key role when students work and learn with team members. According to our quantitative findings, students’ perceived team members’ valuable contributions positively relate to their motivation (H1a) and enjoyment (H1b) from FC-TBL, and individual learning outcomes (H1c). Accordingly, the learners cared about their team members’ value and contributions to the team. If they perceived higher value and contributions of the team members, they are more willing to be devoted to team work as well. Furthermore, they are much happier to share their knowledge of course related materials with team members through FC-TBL. The best thing is that when their perceptions about team members’ valuable contribution are higher, the students eventually learn more effectively.

We also found that perceived motivation (H4a) and enjoyment (H4b) are two mediators when talking about the relationship between perceived team members’ valuable contributions and learning. That is, when students worked on teams, although they strongly cared about other team members’ contributions and value, their
motivation and enjoyment toward FC-TBL were the determinants about learning outcomes. The findings also show that students’ perceived motivation from FC-TBL is positively related to perceived learning (H2a) as well as to perceived enjoyment (H2b). Namely, when students were highly motivated to learn via FC-TBL, they tended to enjoy this flipped learning context, and also, they could learn better. About learners’ motivation, some students praised how the course design is very useful to them, so they had more confidence in their ability to learn in this type of course. They also liked to use the platform and assignments because they felt increasing learning effectiveness. Similarly, we also found that perceived enjoyment from FC-TBL can positively impact learning (H3). That is, when students were more likely to enjoy the team activities in this flipped learning context, they would be able to eventually achieve higher levels of learning.

The aforementioned findings resonated with the theories and empirical studies related to social exchange (Chen, 2013; Emerson, 1976; Jin, Park, & Kim, 2010). FC-TBL was formulated to foster individual student’s learning by contributing to task completion and reciprocating with each other via information sharing and collaboration. Students were required to preview videos relevant to course content to gain basic knowledge of the course before class. As some students stated, they indeed benefit from teamwork in that they usually help each when they have learning problems. The effect of team learning took place when some team members finished the course videos earlier. Afterwards, they discussed the topics and exchanged ideas with one another. They shared and organized opinions, urging the team members to feel the necessity to take responsibility for the assigned teamwork. In addition to the pre-class preparatory work, students needed to help each other during in-class activities and post-class discussion as well. FC-TBL design facilitated students’ teamwork, and also leveled their understanding of textbook chapter content before class, work better in class, and have more interactions and discussions after class.

Conclusions and future research

A flipped classroom shifts the focus of learning from a teacher-centered to a student-centered paradigm (Kim, Kim, Khera, & Getman, 2014; Kuiper, et al., 2015; McLaughlin, et al., 2014). This paradigm is beneficial for active learning, satisfaction, and learning achievement by engaging students (Armbruster, et al., 2009; Michael, 2006; Tune, et al., 2013). In addition, learning in small groups or with team members might be a good approach to promote active and efficient learning (Michaelsen et al., 2002). Grounded in the notion of flipped classrooms and TBL, we developed the FC-TBL learning modules and applied them to HRM courses. Gomez, Wu, and Passerini (2010) found that computer-assisted team-based learning is positively useful to enhance learning experiences. In order to better understand the effects of flipped classrooms on team-based learning, we extended their evaluation framework of CS-TBL to FC-TBL. Our findings indicated that the effectiveness of team-based learning can impact learners’ learning motivation, enjoyment and outcomes in flipped classrooms. Students are more likely to learn better once they perceive that their team members are devoted to team projects. Finally, students’ motivation and enjoyment about FC-TBL are also two critical factors which enhance their learning experience and quality.

Although we tried to incorporate critical factors to investigate students’ team-based learning in this flipped classroom, this study still has a number of limitations. First, the evaluation of the effect of FC-TBL HRM courses on learning outcomes were generally analyzed without eliminating some factors, such as other teachers’ teaching styles, learners’ characteristics, or types of universities. Thus, generalizability might be limited. Second, we interviewed students only to augment the understanding the quantitative results, so the richness and depth of quality qualitative data are limited. Third, more team-based activities can be modified or added to enrich our FC-TBL learning modules, which can better increase learners’ intentions to learn. Future researchers are encouraged to incorporate some individual variables, such as personalities, learning styles, to test whether these characteristics can have impact on learning outcomes. In addition, the present study found that the FC-TBL has positive outcomes in the courses of HRM; future research can conduct it in multiple courses, disciplines or educational levels.

Acknowledgments

The authors would like to thank our anonymous reviewers for their valuable feedback and the Ministry of Science and Technology of Taiwan for its research grant support (MOST 104-2410-H-141-013-MY2 & MOST 105-2410-H-305-075). All errors are exclusively those of the authors.
References


Enfield, J. (2013). Looking at the impact of the flipped classroom model of instruction on undergraduate multimedia students at CSUN. TechTrends, 57(6), 14-27.


Nielsen, L. (2012). Five reasons I’m not flipping over the flipped classroom. Technology & Learning, 32(10), 46.


Flipped Classroom with Problem Based Activities: Exploring Self-regulated Learning in a Programming Language Course

Ünal Çakıroğlu* and Mücahit Öztürk²
1Department of Computer and Instructional Technologies, Faculty of Education, Karadeniz Technical University, Trabzon, Turkey // 2Department of Computer Technologies, Ortaşoy Vocational School, Aksaray University, Aksaray, Turkey // cakiroglu@ktu.edu.tr // mucahitozturk@aksaray.edu.tr

*Corresponding author

ABSTRACT
This study intended to explore the development of self-regulation in a flipped classroom setting. Problem based learning activities were carried out in flipped classrooms to promote self-regulation. A total of 30 undergraduate students from Mechatronic department participated in the study. Self-regulation skills were discussed through students’ and the instructor’s experiences including their opinions and behaviours. Qualitative data was collected with an observation form, discussion messages and interviews with selected participants. As a result, in terms of self-regulated learning, the goal setting and planning, task strategies and help seeking skills of the students were high in the face to face learning designed with problem based activities through flipped classroom model, their goal setting and planning, task strategies and help seeking skills were appeared moderately. In the home sessions, environment structuring, goal setting and planning skills were developed in high level while task strategies, help seeking, time management, monitoring, self-efficacy and self-evaluation skills were moderate and monitoring skills was lower. Consequently, it is hoped that the study may provide some suggestions for using problem based activities in flipped learning.

Keywords
Flipped classroom, Problem based learning, Self-regulated learning

Introduction
In recent years, researchers reported that students in many higher education programs cannot have opportunity for being active namely applying theoretical knowledge on real problems in the classrooms (NMC, 2015). In this sense, problem solving, collaborative group works, self-evaluation, peer tutoring as the active learning strategies become more preferable for making students active in the environment (Kim et al., 2014; McLaughlin et al., 2015). In this sense instead of traditional teaching methods, it is suggested to create learning environments including active learning strategies for problem solving (Barak et al., 2007; Marbach-Ad & Sokolove, 2002).

On the other hand, programming language course is one of the prominent courses in which active learning strategies may be conducted. Since learning programming is generally difficult including programming structures and syntax, critical thinking, and acquire problem solving which are considerably related to problem solving (Kelleher & Pausch, 2005). Thus, programming instructors work on various approaches in order to make students active in learning process. Strayer (2012) asserted that, it is not easy for the teachers to balance presentation and active learning strategies in face to face (F2F) settings. At this point, one such approach gaining popularity in recent years is flipped classroom model (FCM) or flipped learning. This study focuses on teaching programming via FCM by providing problem based activities. Students’ experiences are analysed through understanding how they act during FCM in terms of self-regulation.

Flipped learning
Flipped learning suggests students studying through interactive technologies such as watching videos at home online and preparing themselves to apply active learning strategies in the classroom (Bergmann & Sams, 2012; Herreid & Schiller, 2013; Roach, 2014). Students take notes and prepare questions about the topics with the theoretical knowledge at home. They electronically share their questions with the teacher at home and receive instant feedback in F2F settings (Bergmann & Sams, 2012). Teacher classifies the questions regarding the qualifications of the students and can prepare for discussions and feedback in the classroom (Fulton, 2012). Teacher and student roles in FCM are briefly summarized in Figure1.

Since, FCM includes both F2F and online settings, the focuses of the research studies are somewhat different from typical F2F settings. In this regard, academic achievement (Al-Zahrani, 2015; Giannakos et al., 2015; Wanner & Palmer, 2015) and the attitude (Davies et al., 2013; Roach, 2014) were frequently investigated in FCM studies. In this regard, FCM was found to promote the participation of the students, enhance their problem solving abilities in group work (Berrett, 2012) and give opportunity to the teacher for individual guidance (Mok,
Some other studies analysed students’ perceptions, motivation and their self-efficacies (Findlay-Thompson et al., 2014; Kong, 2014; Mason et al., 2013).

On the other hand, self-regulation is another issue addressed in online learning environments (Tabak & Nguyen, 2013; Yükselturk & Bulut, 2009). Self-regulation is generally considered from the point of revealing students’ control skills about what, when and how to study (Cunningham & Billingsley, 2002). In this sense, self-regulation has a remarkable place in FCM whereas majority of learning in FCM occurs in online learning settings.

Study framework

This study focuses on self-regulation in the problem based activities. It considerably goes around the aspects of FCM, problem based learning and self-regulation including the framework summarised in Figure 2.

Problem solving contributes to develop real life problem solving skills, creativity, productivity and flexibility (Lazakidou & Retalis, 2010). It also comprises student engagement, critical thinking, self-directed learning, authentic learning, group skill development (Klegeris & Hurren, 2011; Capon & Kuhn, 2004). In problem based learning (PBL), generally unstructured problem cases are provided for students. Students define the problem, collect data, and conduct group discussions for analysing data and produce solutions in a collaborative work (Pease & Kuhn, 2011). Roach (2014) pointed out that, when conducting FCM, if the tasks at school are organized as problem based activities, students may benefit from the potential of collaboration. Because, students study on problems at home through online platforms and they perform PBL activities, projects or homework in F2F sessions. As a matter of fact, some researches asserted that FCM may provide quality learning outcomes in which problem solving skills are taken into consideration (Berrett, 2012; Missildine et al., 2013).

On the other hand, Bland (2006) asserts that self-regulation is required for meaningful learning where the students densely have their own learning control in PBL environments. SRL is metacognitive, motivational,

Evaluating self-regulated learning


Zimmerman and Martinez-Pons (1990) consider SRL strategies as self-evaluation, organization and transformation, goal setting and planning, seeking information, keeping records, monitoring, environment structuring, self-consequences, rehearsing and memorizing, seeking social assistance and reviewing records. In addition, Barnard et al. (2009) pointed out some specific characteristics for online learning environments. Since FCM includes both in and out of school activities, this study drew on the framework suggested by Barnard et al. (2009), Pintrich et al. (1993) and Zimmerman and Pons (1990).

Aim of the study

The aim of this study is to explore how the SRL skills of the students develop in a flipped learning environment in which problem based activities were used. The research question “To what extend did the students develop SRL skills in the flipped classroom with problem based learning?” guided to the research.

Method

This case study is based on students’ and the instructor’s perceptions and experiences in an introductory programming course. Development of SRL skills were also observed during ten weeks period.

Participants

Participants were 30 students (25 male, 5 female) enrolled in Mechatronics department of a public university in Turkey. They received introductory programming course for the first time. They had basic knowledge to follow videos on the LMS but they had no experience about flipped learning.

Procedure

One of the researchers was instructor and the other was observer in the classroom. In the first four weeks, Algorithm topic was taught in F2F instruction. The students were divided into six groups according to Algorithm test scores at the end of the four weeks. In the following five weeks, the main issues in C++ programming language were delivered via FCM. Students followed the videos weekly at home and discussed on various problems. In the F2F session, students in the groups collaboratively provided solutions for the problems and the instructor guided the students about the problems. Generally; at the end of the lessons, a selected student presented the solution of the problem in the classroom. The implementation is summarised in Figure 3.
Online learning platform

Online Learning Platform (OLP) has three main modules: Course, Test and Task developed by the researchers.

Course module

The course videos were prepared through capturing display images through the Dev C++ Editor and two videos were weekly uploaded in this module. Students asked questions to the instructor and their peers by using comment section. At the end of the videos, each group was given problems about the related topic. Group members discussed about the problems before meeting in F2F lesson. Course module is shown in Figure 4.
Test module

Students answer questions after watching videos in this module. Anyone who failed at the end of the test was directed to watch the videos again. The test statistics including students’ achievements in the tests and the number of the tests they answered were kept by OLP. Instructor shared the statistics in F2F sessions.

Task module

Problems given at the end of the videos was separately shared in the Task module for each week. In addition, a Facebook group or emails were used to get in touch with the instructor and students each other.

Three modules in this study is summarised in Figure 5.

![Figure 5. Online Learning Platform](image)

Instruments

In this study, SRL skills were taken into consideration in F2F and in online environments. Accordingly, the items were referred from various scales to cover all the SRL skills in FCM. It was noticed that factors developed by Zimmerman and Pons (1986), and Pintrich et al. (1993) were generally employed in similar studies (Toering et al., 2012; Barnard-Brak et al., 2010; Cho & Jonassen, 2009). In this sense, for F2F sessions items in the scales of (Zimmerman & Pons (1986) and Pintrich et al. (1993)) were used. Also definitions of Barnard et al. (2009) were referred for SRL in online video sessions. The SRL dimensions referred in the instrument were briefly shown in Figure 6.

Figure 6 indicates that, “Environment structuring, Goal setting and planning, Help seeking” are common dimensions in three definitions. “Time management, Rehearsal, Organization, Monitoring, and Self-evaluation” are common at least in two definitions. Although “Transformation, Memorizing, Self-efficacy, Task strategies, Elaboration, Regulating” are specific factors for three scales, they also have some common characteristics in terms of their content. For instance, Pintrich and Schrauben (1992) consider the factors of “Rehearsal, Organization, Elaboration, Regulating, Memorizing, Transformation” under the topic of “Learning Strategies” while Barnard et al. (2009) analyzed them as a different factor as the “Task Strategies.” Also, “Self-efficacy” was analyzed as a separate factor. Considering the common and prominent dimensions, the analysis was directed through the scope of eight sub-dimensions: “Environment structuring, Goal setting and Planning, Task strategies, Help seeking, Time management, Monitoring, Self-efficacy and Self-evaluation.” The following instruments were used to collect data in order to explore the developments of SRL skills.

Observation Form: In order to notice the development of SRL skills, an observation form was developed and used by one of the researchers in the process. The form consisted of the factors “Environment structuring, Goal setting and planning, Help seeking, Task Strategies, Monitoring, Time Management, Self-efficacy, and Self-evaluation.” While addressing the items in the form, the observations about the behaviors, perspectives, and interactions were coded as positive or negative. For example, the items such as “I allocate extra studying time for my online courses because I know it is time-demanding,” “I try to schedule the same time every day or every
week to study for my online courses, and I observe the schedule were negatively or positively recorded in the observation form for the factor of “Time Management” depending on the status of the students in terms of conducting those attitudes. Researcher filled the form for both F2F and home sessions.

**Figure 6. Self-regulated Learning Skills**

**Interviews:** In order to explain the observations in detail, interviews were carried out with six selected participants. One student from each group was selected. Interviews were conducted by one researcher through a structured form including 8 factors of the SRL skills. The interviewees were selected because they had no absence throughout the academic year, participated in all of the exams, and they were volunteered to participate.

**OLP Logs and Social Network Messages:** For each student, the number of videos followed, the number of comments, and students’ scores from the online tests were recorded in OLP Logs. A Facebook group was employed to provide a faster communication among the students. Students’ SRL skills at home sessions were evaluated within the number of watching videos, messages on OLP or Facebook, the period of time they spent on OLP.

**Analysis**

The analysis of the process in the research is presented in Figure 7.

**Figure 7. Data analysis**

While determining the SRL skills in each week, eight factors of SRL skills and the frequency of the behaviors which were considered “positive” were taken as a basis. The observed frequencies and their durations were calculated and classified. The scores between “0 and 60” were regarded “Low,” those between 60 and 85 were regarded as “Medium” and the scores between 85 and 100 were assessed as “High.” Some other studies also provided similar classifications through the frequencies of students’ behaviors for interpreting data on observation form (Burgess et al., 2010). For example, the weekly frequency was observed through adding the
number of the students whose perspective was regarded “positive” for the factor of “Environment Structuring” in the Observation form.

The data obtained from the interview form was classified into the themes which are extracted due to eight factors of SRL skills. The observation data were elaborated and associated with the interview data.

Results

Results are organized as reporting SRL developments first, and then reporting the contributions of problem based activities.

Students’ SRL skills in the flipped learning

OLP logs and Facebook messages are evaluated together, and the developments of students SRL skills were outlined in Table1. The frequency is presented for the F2F and home sessions.

<table>
<thead>
<tr>
<th>Table1. Development of students’ SRL skills</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>5 Week 6 Week 7 Week 8 Week 9 Week Average</td>
</tr>
<tr>
<td>F2F Home F2F Home F2F Home F2F Home F2F Home</td>
</tr>
<tr>
<td>Environment Structuring</td>
</tr>
<tr>
<td>M M M H M H M M H H M H</td>
</tr>
<tr>
<td>Goal Setting and Planning</td>
</tr>
<tr>
<td>M M H H H H M M H H H H</td>
</tr>
<tr>
<td>Task Strategies</td>
</tr>
<tr>
<td>M L H M H M L H M H M H M</td>
</tr>
<tr>
<td>Help Seeking</td>
</tr>
<tr>
<td>M M H M L H L H M H M M</td>
</tr>
<tr>
<td>Time Management</td>
</tr>
<tr>
<td>M L M H M L M L H M M M</td>
</tr>
<tr>
<td>Monitoring</td>
</tr>
<tr>
<td>M L M H M L M L M M L M</td>
</tr>
<tr>
<td>Self-efficacy</td>
</tr>
<tr>
<td>L L M M M M M M M</td>
</tr>
<tr>
<td>Self-evaluation</td>
</tr>
<tr>
<td>M L M M M M M M M</td>
</tr>
</tbody>
</table>

Low: Self-regulated Learned Skills, 0-60% of Students
Medium: Self-regulated Learned Skills, 60-85% of Students
High: Self-regulated Learned Skills, 85-100% of Students

The SRL skills developments were illustrated with frequencies which were determined through the data from OLP discussions, tests, Facebook messages and observation notes. For example, in Environment structuring dimension, the number of students who followed the videos were gathered from OLP reports, the number of students participated to collaborative activities or the number of students provided some arrangements at home were gathered via observation notes. For 5-9 weeks implementation, the average number of students for all weeks who showed indicators about SRL skills were determined and outlined in Figure 8.

Figure 8 indicates that in the implementation, skills about goal setting and planning dimensions were quite high. The skills of environment structuring, task strategies, help seeking, time management, self-efficacy and self-evaluation were exhibited by most of the students. However, the number of students who exhibited monitoring skills was relatively lower than other skills.

It was observed that majority of the students developed their “Environment structuring” skills in F2F and home sessions. The students regularly watched the videos through using the OLP at home for 5-9 weeks. They generally participated to the collaborative group activities and organized their own learning in F2F setting. In this sense, S2 stated that “I attended the face to face courses after watching the videos regularly at home ...” It was also observed that the students came together with their group-mates and started to study when they entered in the F2F sessions. They conducted various arrangements at home in order to study more effectively while watching the videos in the home. S4 addressed that “…The noisy environments decrease my motivation I watched the videos related to the courses at home at peace and quiet. …” Since, in the implementation, the collaborative group works were conducted in F2F environment. It was observed that the students employed different approaches related to the collaborative group activities. Also, some students in groups preferred individual activities. Some students in Group-3 and Group-5 focused on the problem case instead of cooperative group activities in F2F sessions.
Figure 8. Average Number of Students who showed SRL skills

When goal setting and planning factor was analyzed, it was seen that SRL skills developed most commonly among the students. It was also observed that the students had necessary information about the problem case given to their group and they conducted discussions accordingly. They prepared for the given problem cases given to them. S4 expressed his opinion as follows; “After I had watched the videos, I knew what to do during the face to face lesson upon the problem cases is given. We conducted discussions with our group-mates related to the homework through social media...”. It was found that the students regularly studied on the contents of the courses they have to learn at home through the OLP environment and they came to F2F sessions after a preparation.

It was seen that majority of the students developed “Task strategies” skills in F2F and Home sessions. It was observed the students repeatedly watched the videos in order to understand the concepts in OLP environments at home. They took notes and applied the applications of the programming codes through their computers and gave their consideration to the problem case given to them. S5 explained the case as following; “…I watch the videos of the course, I tried to code after I watch. When I fail I watch the videos again ....” It was seen that the students repeatedly watched the videos when required while they were studying on the problem case in F2F sessions and they analyzed the samples of coding in OLP. In the cooperative groups, the students primarily studied on the problems individually. Afterwards, they presented each other their suggestions for the solutions in discussions. Thus, they tried to reach to a conclusion.

It was addressed that majority of the students developed “Help seeking” skills in F2F and Home sessions. Students identified that they received support from their peers or instructors related to the content of the course learning environments at home. At home, they directed their questions through comments section of the OLP or through Facebook group in the weekly periods. Moreover, they shared their solutions for the tasks in weekly groups at the end of each F2F lesson in the Facebook group. They asked questions generally to the instructor through OLP and they get in touch with their group mates through private messages. S6 stated that; “…Initially, I asked the questions about the issues I failed in comprehending to my peers. I asked the instructors if I can’t find the correct answer...”. Also, students got help from their peers, web and the instructor respectively in the F2F sessions. The instructor didn’t directly give answers to the questions asked by the students, he guided the students to do researches and discuss the problems with their peers in their group. Moreover, the students were also helped each other.

Majority of the students also developed “Time Management” skills in F2F and Home sessions. The time management of the students is seen as the time which the students allocated especially to watch the videos regularly OLP environment at home. S2 stated; “I regularly watched the videos every week. I allocated about 30-45 minutes a week in average...”. S1 emphasized that they had time left for practicing and added; “…We obtained some information about the issue through videos and come to the classroom. I only deal with the problem cases given to us. Thus, we have free time to make exercises related to the issue we have just learned...”. It was observed that the students conducted the task given to them on time in the F2F environment and completed them within the related lesson.

When the monitoring factor is analyzed, it was found that the SRL skill developed least in the OLP environment at home. It was observed that the students didn’t study much on the problem case given to their groups in the
OLP environment at home. S5 stated that "... I conducted the application of the codes given in the videos. However, I didn’t study the problems at home..." The students studied the problem cases in the F2F sessions. However, it was found that they didn’t study on the problem cases of other groups and the solutions for them. Thus, it was concluded that they didn’t prefer watching their own learning levels through studying on the different problem cases. The students studied on the codes which were explained in videos and they mostly did the homework related to the application of the codes in different occasions, mostly in F2F environment.

Majority of the students also developed their “Self-efficacy” skills in F2F and Home sessions. It was found that the students experienced no problem in learning the contents in the OLP environment at home. S1 emphasized that, “...I generally understand the issues taught in the videos. If I encounter problems, I watch them again....” while S6 emphasized that “... I can make codifications given in the videos by myself...” On the other hand, the students mostly had difficulty when level of difficulty of the given problem cases increased. S2 stated that “... I had difficulty in doing homework when he level of difficulty for the issues increased in the course of time. For that reason I tried to do my homework after I received help from my friends or the instructor...”. It was also observed that the students worked with group-mates since they faced difficulties in studying individually in F2F sessions.

In addition, majority of the students developed “Self-evaluation” skills in F2F and home sessions. The students wrote down whatever they perceived from the content and their questions on the comment section under the videos which they regularly used and through the encouraging of the instructor OLP environment at home. Moreover, they regularly solved the online tests in OLP environment. When they failed, they solved online tests after they removed their deficiencies. S6 stated that: “After I solved the questions in the test module, I was directed to watch the videos again when I got a score under a definite level. The harder the issues got the more I had difficulty in accomplishing the test only once...” In the F2F sessions, the students shared their solutions with other groups at the end of the lesson. Thus, it was observed that the students were able to analyze the solution for the different problems.

Overall, the data obtained from the observation notes and interviews indicate that, the levels of “Goal setting and planning, Task strategies, Help Seeking” skills were considerably high in F2F learning environments among the students. Also, their “Environment structuring, Time management, Self-efficacy, Self-evaluation, and Monitoring” skills may be considered at medium level. In home sessions, while “Goal setting and planning, Environment structuring” skills were high, “Task strategies, Help Seeking, Time management, Self-efficacy, Self-evaluation” skills were medium and the “Monitoring” level was low.

**Discussion and conclusion**

In this study, the problem based learning in the flipped classroom was analyzed within the SRL framework. In the study, problem solving activities were given prominence in F2F setting in FCM which contains problem based learning when compared to other learning activities. At this point, the Course module in OLP became prominent so that the students watch the videos and they wrote down their questions and whatever they understood. Students used Test module for self-evaluation. Using the Task module, they acquired information about the problems given to them weekly for the F2F learning environment. Students worked collaboratively in F2F learning environment related to the problem event given to their group.

The occasion of executing the Environment structuring skills of the students through the problem based FCM was higher at home when compared to F2F learning environment. Almost all the students watched the course videos regularly. They provided some arrangements in their learning environment in order to benefit efficiently from the videos. It was found that the students planned their own learning activities through Problem based FCM, they motivated themselves for studying and arranged the learning environment as suggested by Artino and Stephens (2009). Most of them confirmed that they worked collaboratively in the F2F learning environment. Especially, successful students preferred to work individually. But when they work collaboratively, they took responsibility in the group activities more than the other students. Students having difficulty in understanding the content preferred collaborative environment, because they received help from peers and motivated each other.

According to Wanner and Palmer (2015), the students mainly prefer collaborative learning activities in F2F environment although they conduct the learning activities through online platforms in FCM. Similar to this idea; students were responsible to organize the learning environments at home and the collaborative environment in the classroom was formed through the guidance of the instructor. Thus, the difference between the Environment structuring skills of the students in the F2F and home environment may be resulted from those cases.
Setting up the objectives in problem based environments is regarded as one of the indicators of SRL skills among the students (Loyens et al., 2008). In this study, the executing the skills of Goal setting and planning among the students were found to be high in both F2F and at home. The students employed OLP weekly at home systematically and they planned their works. They had information about the content and objective of the videos in OLP they use at home. On the other hand, they were aware of the problems case given to them for F2F learning environment. Thus, the characteristics of the learning environment somehow affected the SRL skills of the students. As a matter of fact, the students came to F2F sessions knowing what they would do in the F2F environment although they didn’t work on the problem case much at home. So, problems provided for students may be concluded to contribute on the point of determining and planning the objectives of the students.

It was found that the students had higher rates for Task strategies in the F2F environment than watching the videos at home. The results indicated that the students encountered difficulties in explaining the meaning of the given problems when they could not provide adequate effort for the theoretical knowledge at home. As seen in similar studies, students in this study tried to overcome the deficiencies at this point through the supports from the instructor and peers through discussing and repeating in the F2F setting (Touchton, 2015; Artino & Stephens, 2009). The opportunity to watch the videos more than once provided much contribution when the students needed to associate with the previous issues. Upon the students learn the content to a certain extent in the F2F setting; they were able to take an active role in learning in the classroom as seen in some studies which conducted similar applications related to problems in the classroom (Love et al., 2014).

In the dimension of help for seeking skills, students mostly preferred F2F learning environment to receive help on the topics they had difficulty. They asked questions to other members and the instructor through OLP or Facebook. As a matter of fact, the students asked for help more in the classroom when they failed in understanding the content at home. The instructor conducted applications for learning at home through FCM provided contributions to the students at the point of learning through solving problems. The role of the instructor was similar to another study who effectively applied the problem solving through FCM and collaborative activities (Al-Zahrani, 2015).

Time management skills of the students were developed similarly both in F2F and learning at home. Since the students came to class having certain knowledge, they gained time to conduct applications in F2F setting and it provided opportunities to use the time at home efficiently. To that end, it helped in reducing the time which passes with teaching in classroom as seen in the FCM; thus, a top-level learning was provided (Steed, 2012).

Monitoring skills among the students were lower at home. A few students were able to apply the knowledge they obtained personally to different cases. Very few students could apply the knowledge they learned at home into a different case. The students didn’t need it since they would study on the problem case given to their groups in the F2F learning environments. In some cases, it was pointed out that the students were able to learn at the application level and higher in F2F learning environment (Touchton, 2015; Seamen & Gaines 2013).

Through problem based activities majority of the students felt qualified within the context. However, it is also recognized that self-efficacy of the students decreased during the weeks when they had difficulty in perceiving the issues. In that case, they can receive help in the F2F setting from their peers and instructors. The students were generally able to perform the tasks in the activities about the topics which are explained in the videos. Although they could not find the absolute solutions for the given programming problems, they were able to bring proposals forward. In this sense, problem based activities has positive contributions to the self-efficacy among the students (Wiginton, 2013).

The students measured their level of knowledge regularly using the test module in the OLP at home. According to the result coming from this result, they watched the videos again or thought about the given problem case. As Artino and Stephens (2009) pointed out that the students gained opportunities to evaluate themselves, seeing their deficiencies and making regulations in this way, students in this study forwarded the sections they couldn’t understood to the instructor of their peers in OLP, Facebook or F2F learning environment.

Some researchers argue that SRL skills were in relation with problem solving skills (Veenman, 2006; Veenman & Beishuizen, 2004). According to Lazakidou and Retalis (2010), problem based learning is a process which SRL strategies are implemented to achieve solution for the given problem. From this point of view, it is thought that SRL skills may be developed since PBL and SRL strategies are provided together. Generally, it is considered difficult for students to develop their SRL skills in the online learning settings (Sun & Rueda, 2012), so FCM may enable active learning strategies through its hybrid structure. Thus, the results of the research provided an insight that students SRL skills may be developed through Problem based FCM.
It was found that the Goal setting and planning, Task strategies and Help seeking skills were high in the F2F setting. In addition, their Goal setting and planning, Task strategies and Help seeking was appeared moderately. In the home sessions, students skills of Environment structuring, Goal setting and planning were high level while Task strategies, Help seeking, Time management, Monitoring, Self-efficacy and Self-evaluation skills were moderate and Monitoring skill was lower. A conclusion may be drawn that the problem based activities in FCM could provide positive contributions to the students’ Self-regulated Skills. On the other hand, the opportunity of repetition at home has enabled students to perform applications through collaborative group activities in the F2F settings and facilitate learning programming which require problem solving skills.

Consequently, this study is hoped to provide an insight for using problem based activities in flipped learning. Although the study is limited to the use of problem solving and collaborative group activities among the active learning strategies in FCM; for future studies, other active learning strategies may be taken under consideration. Moreover, monitoring skills which is one of the crucial dimension in SRL was found low in this study and it may be specifically analyzed in FCM. Furthermore, the relationship between SRL, PBL and FCM model in different disciplines may be examined.

References


Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. Eugene, OR: International Society for Technology in Education.


Veenman, M. V., & Beishuizen, J. J. (2004). Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint. *Learning and Instruction, 14*(6), 621-640.


