Educational Technology & Society
An International Journal

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

- Educators aim to use technology to enhance individual learning as well as to achieve widespread education and expect the technology to blend with their individual approach to instruction. However, most educators are not fully aware of the benefits that may be obtained by proactively harnessing the available technologies and how they might be able to influence further developments through systematic feedback and suggestions.

- Educational system developers and artificial intelligence (AI) researchers are sometimes unaware of the needs and requirements of typical teachers, with a possible exception of those in the computer science domain. In transferring the notion of a “user” from the human-computer interaction studies and assigning it to the “student,” the educator’s role as the “implementer/manager/user” of the technology has been forgotten.

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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

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


Founding Editor
Kinsuk, University of North Texas, USA.

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

Advisory Board
Ignacio Aedo, Universidad Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Nicholas Chang, Indiana University South Bend, USA; Yann 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; Tong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; Richard Isaac, National Institute of Information Technology, India; Mohamed Jenni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasios Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Rob Koper, Open University of the Netherlands, The Netherlands; Jimmy Ho Man Lee, The Chinese University of Hong Kong, Hong Kong; Rudy Leolouche, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McCreath, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Vaux University, Sweden; Riichiro Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoshi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J. Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan; Yu-Mei Wang, University of Alabama at Birmingham, USA; Ashok Patel, CAL Research & Software Engineering Centre, UK; Reinhard Oppermann, Fraunhofer Institut Angewandte Informationstechnik, Germany; Vladimir A Fonichkov, K. E. Tsolovskiy Russian State Tech Univ, Russia; Olga S Fomichova, Studio “Culture, Ecology, and Foreign Languages,” Russia; Piet Kommers, University of Twente, The Netherlands; Chul-Hwan Lee, Inchon National University of Education, Korea; Brent Muirhead, University of Phoenix Online, USA; Erkki Sutinen, University of Joensuu, Finland; Vladimir Uskov, Bradley University, USA.

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-Chich 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.
Supporting Organizations

- Center for Cognition and Gesture-based Computing, National Sun Yat-sen University, Taiwan
- Taiwan E-Learning and Digital Content Association (TELDCA), Taiwan
- Information Technologies Institute, Centre for Research & Technology Hellas (C.R.T.H), Greece
- University of Piraeus, Greece

Subscription Prices and Ordering Information

For subscription information, please contact the editors at ets-editors@ifets.info.

Advertisements

Educational Technology & Society accepts advertisement of products and services of direct interest and usefulness to the readers of the journal, those involved in education and educational technology. Contact the editors at ets-editors@ifets.info.

Abstracting and Indexing


Guidelines for authors

Submissions are invited in the following categories:

- Peer reviewed publications: Full length articles (4000 - 7000 words)
- Book reviews
- Software reviews
- Website reviews

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

For detailed information on how to format your submissions, please see:

http://www.ifets.info/guide.php

Submission procedure

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

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

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

Please provide following details with each submission:

- Author(s) full name(s) including title(s),
- Name of corresponding author,
- Organisations,
- Full contact details of ALL authors including email address, postal address, telephone and fax numbers.

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

**Full Length Articles**

1. **Game On! Students’ Perceptions of Gamified Learning**  
   Patrick Buckley, Elaine Doyle and Shane Doyle  
   1–10

2. **Exploring the Effects of Online Academic Help-Seeking and Flipped Learning on Improving Students’ Learning**  
   Wen-Li Chyr, Pei-Di Shen, Yi-Chun Chiang, Jau-Bi Lin and Chia-Wen Tsai  
   11–23

3. **Web-Based System for Adaptable Rubrics: Case Study on CAD Assessment**  
   Pedro Company, Manuel Contero, Jeffrey Otey, Jorge D. Camba, María-Jesús Agost and David Pérez-López  
   24–41

4. **Differences in Views of School Principals and Teachers regarding Technology Integration**  
   Magdalena Claro, Miguel Nussbaum, Ximena López and Victoria Contardo  
   42–53

5. **Synchronized Pair Configuration in Virtualization-Based Lab for Learning Computer Networks**  
   Chaknarin Kongcharoen, Wuyin Hwang and Gheorghita Ghinea  
   54–68

6. **Technology-Enhanced Peer Review: Benefits and Implications of Providing Multiple Reviews**  
   Pantelis M. Papadopoulos, Thomas D. Lagkas and Stavros N. Demetriades  
   69–81

7. **Effects of a Structured Resource-based Web Issue-Quest Approach on Students’ Learning Performances in Computer Programming Courses**  
   Ting-Chia Hsu and Gwo-Jen Hwang  
   82–94

8. **Student Engagement in Long-Term Collaborative EFL Storytelling Activities: An Analysis of Learners with English Proficiency Differences**  
   Yun-Yin Huang, Chen-Chung Liu, Yu Wang, Chin-Chung Tsai and Hung-Ming Lin  
   95–109

9. **The Influences of the 2D Image-Based Augmented Reality and Virtual Reality on Student Learning**  
   Hsin-Hun Liou, Stephen J. H. Yang, Sherry Y. Chen and Wernhuaar Tarn  
   110–121

10. **An Investigation of Technological Pedagogical Content Knowledge, Self-Confidence, and Perception of Pre-Service Middle School Mathematics Teachers towards Instructional Technologies**  
    Ilhan Karatas, Mutlu Piskin Tunc, Nurbanu Yilmaz and Galzade Karaci  
    122–132

11. **Social and Collaborative Interactions for Educational Content Enrichment in ULEs**  
    Rafael D. Araújo, Taffarel Brant-Ribeiro, Igor E. S. Mendonça, Miller M. Mendes, Fabiano A. Dorçan and Renan G. Cattelan  
    133–144

12. **Students’ Metacognition and Cognitive Style and Their Effect on Cognitive Load and Learning Achievement**  
    Omar López-Vargas, Jaime Ibáñez-Ibáñez and Oswaldo Racines-Prada  
    145–157

13. **Improving Learning Analytics – Combining Observational and Self-Report Data on Student Learning**  
    Robert A. Ellis, Feifei Han and Abelardo Pardo  
    158–169

14. **Metacognitive Support Accelerates Computer Assisted Learning for Novice Programmers**  
    Siti Nurulain Mohd Rum and Maizatul Akmar Ismail  
    170–181

15. **Investigation of Continuous Assessment of Correctness in Introductory Programming**  
    Deller James Ferreira, Hebert Coelho da Silva, Tatiane F. N. Melo and Ana Paula Ambrósio  
    182–194

16. **A Study of Supplementing Conventional Business Education with Digital Games**  
    Abida Ellahi, Bilal Zaka and Fahd Sultan  
    195–206

17. **Using Mobile Learning to Support Students’ Understanding in Geometry: A Design-Based Research Study**  
    Helen Crompton  
    207–219

18. **Investigating the Period of Switching Roles in Pair Programming in a Primary School**  
    Baichang Zhong, Qiyun Wang, Jie Chen and Yi Li  
    220–233
Game On! Students’ Perceptions of Gamified Learning

Patrick Buckley¹, Elaine Doyle¹* and Shane Doyle²

¹Kemmy Business School, University of Limerick, Ireland // ²MCCP, The Rear Warehouse, Dublin, Ireland // Patrick.buckley@ul.ie // Elaine.doyle@ul.ie // shane@mccp.ie

*Corresponding author

(Submitted November 13, 2015; Revised March 7, 2016; Accepted April 25, 2016)

ABSTRACT

Gamification is presented in the literature as a pedagogical innovation that may increase student engagement and enhance learning. This study explores students’ perceptions of a gamified learning intervention deployed in a large undergraduate module and a small postgraduate module. Given the dearth of previous empirical work, an exploratory approach was used. Focus groups were carried out to develop a nuanced understanding of the students’ perceptions of a gamified learning environment. Six themes emerged: impact on learning outcomes, motivation, the importance of the stakes, group dynamics, gender and the challenges gamified learning activities present. The paper contributes by evaluating students’ perceptions of the effectiveness of gamification, providing guidelines for other practitioners deploying gamified learning interventions and identifying outstanding issues and questions that require further research.

Keywords
Gamification, Student motivation, Student engagement, Active learning, Millennials

Introduction

The beginning of the 21st century has seen several macro-level trends challenge the traditional higher education model (Doyle, Buckley, & Carroll, 2013). Most notable is the arrival of a new generational group (Elam, Stratton, & Gibson, 2007; Howe & Strauss, 2000; Howe & Strauss, 2003). Referred to as “Millennials” or “Generation Y,” they are widely seen as being the first “digital natives,” exposed to information technology from birth. Other far-reaching changes have coincided with the arrival of this new cohort. Massification, a concept which refers to the enrolment of students beyond the levels required to repopulate academia and certain other high status professions, continues apace (Cornuel, 2007). This inevitably leads to larger and more diverse classes, with far more variance in student background and ability. University financing models are also changing (Altbach, Reisberg, & Rumbley, 2009). Higher level institutions have adapted to the changing financial environment by developing new revenue streams, including fee paying postgraduate and executive education courses, and by recruiting fee paying students from foreign jurisdictions, amplifying the negative effects of massification.

One concern linked to these trends is that student engagement and motivation is declining. The symptoms are well documented. Lecturers report declining class attendance (Massingham & Herrington, 2006) coupled with difficulties in prompting interaction and discussion (Race, 2010). More pernicious problems such as plagiarism and cheating are reported to be rising (Flint, Clegg, & Macdonald, 2006). In response, educators are striving to develop innovative teaching practices to capture and retain the attention of students, particularly millennials. One approach, and the subject of growing interest, is gamification. This involves “using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems” (Kapp, 2012, p. 10). More generally, gamification is the term applied to a set of motivational triggers, such as rewards and competition, which are traditionally associated with games. While the application of gamification in education is still an emerging trend (Dicheva, Dichev, Agre, & Angelova, 2015), its proponents suggest that it can be employed to enhance student engagement and prompt learning.

However, due to the dearth of empirical research in this area, little is known about how gamification impacts on students’ motivation to engage and learn. For example, in what contexts does gamification have most impact on engagement? Do gamified learning activities work better in large or small classes? Does the level of specialism of the relevant degree make any difference to the effectiveness of gamification in education? What challenges or difficulties do gamified learning environments present for students? When attempting to influence behaviour, a comprehensive understanding of the target group’s perceptions of the relevant issue is necessary (Gullifer & Tyson, 2010).

To this end, a gamified learning intervention was deployed within two university modules taken by different student cohorts during the same semester. The first was a large group of undergraduate students while the second...
was a small postgraduate course. Focus groups were carried out with each cohort to develop a nuanced understanding of their perceptions of a gamified learning environment across different class sizes, level of education and subject specialism. Our paper makes three contributions. First, we assess students’ perceptions of the effectiveness of gamification as a pedagogical technique. Second, we provide other educators with guidance on how gamification can be integrated into curricula. Finally, we use the data to identify outstanding issues and questions that must be fully addressed before gamification can be considered a mature pedagogical methodology. In the following section, we review the relevant literature and identify the attributes of gamified activities. We then describe our research design and data collection method. Our findings are presented before the paper concludes with a discussion of our results and suggestions.

Literature review

Games provoke powerful emotional responses, such as curiosity, frustration and joy (Kim, 2012). Gamification involves using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems (Kapp, 2012, p. 10). It is often closely identified with computer games. However, gamification does not necessarily involve the use of either an actual game or information technology. Instead, it involves the integration of design elements or activity patterns traditionally found in games into non-game contexts (Simões, Díaz Redondo, & Fernández Vilas, 2013).

The literature identifies a number of common attributes of gamified activities. First, individuals receive rewards for achieving goals or overcoming obstacles. This is operationalized in a variety of ways and may include badges, prizes and levelling systems (Glover, 2013). Badges are used to publicly demonstrate that the player has achieved a particular game objective. Prizes fulfil a similar role, but provide the player with a reward extrinsic to the game. Levelling systems are intrinsic to games, and provide players with increased status, access or power within the game environment (Zichermann & Cunningham, 2011).

Levelling systems are also a key enabler of the rapid feedback cycles that constitute the second feature of gamified activities (Lee & Hammer, 2011). Game playing is associated with trial, error, failure and eventual success through practice, experience, reflection and learning. A key objective is not to forbid failure, but develop a positive relationship with it. Failure is not seen as an end, but as a step on the journey to mastery. Gamified learning interventions seek to maintain a positive relationship with failure by creating rapid feedback cycles and keeping the stakes for individual learning episodes low (Lee & Hammer, 2011). Levelling systems, sometimes referred to as progression stairs, mediate these feedback cycles by ensuring that players interact with challenges and other players at a level commensurate with their own competence (Zichermann & Cunningham, 2011).

Both traditional and video games have objective, specific rules (Smith-Robbins, 2011). Salen and Zimmerman (2004) define a game as “a system in which players engage in an artificial conflict defined by rules that results in a quantifiable outcome.” (p. 81). Rules structure the activity, and place clear limits on the actions players can take. For example, in poker, rules forbid players from looking at other players’ cards. Taken as a whole, the rules define the scope of the game.

A fourth distinguishing characteristic is the explicit use of competition as a motivational tool (Nicholson, 2012). Competition may be generated at an individual level through progress tracking, whereby movement towards an overall objective is mapped by a sequence of intermediate goals to be achieved. It may also be social, with individuals competing against each other to achieve the highest score. This is often operationalized in the form of a leader board, which ranks from first to last in terms of performance (Deterding, Dixon, Khaled & Nacke, 2011). These ranking systems serve as a source of motivation because participants see their efforts publicly and instantly recognised (Domínguez et al., 2013).

In successful gamified activities, the attributes listed previously are not disparate elements, but must be knitted together to create a cohesive whole (McGonigal, 2011). For example, the success of competition as a motivational tool depends on whether the reward system is perceived as credible, transparent, challenging and fair. A transparent reward system depends upon having fair and transparent rules, and so on. Such analysis can be continued ad infinitum, and this web of interdependencies means that designing effective gamified activities is challenging (Lee & Hammer, 2011).

The power of gamification as a motivational tool has led to calls for educators to consider how the concept might be applied in educational contexts (Domínguez et al., 2013; Lee & Hammer, 2011). A literature review of empirical studies on gamification carried out in 2014 concluded that gamification has positive impacts, however
its effects are greatly dependent on the context in which the gamification is implemented and the profile of the users (Hamari, Koivisto & Sarja, 2014). Gamification has also been criticised within the education literature (see, for example, Glover, 2013). At a philosophical level, one concern is that it may reduce the internal motivation that the user has for the activity in question by replacing internal motivation with external motivation (Nicholson, 2012). The desirability and utility of competition in educational contexts is another open research question (see, for example, Deardon, 1972, and Rich, 1998). Another concern is that gamification may encourage addictive or compulsive behaviour among people with certain personality types (Zichermann & Cunningham, 2011).

More specifically, there have been many calls in the literature for a more nuanced research agenda which examines gamification at a higher level of resolution. Hamari et al. (2014) call for research to investigate the role of context and the qualities and attributes of users when considering gamified learning interventions. Miller, Cafazzo and Seto (2014) echo this call for research investigating how the characteristics of users impact on the effectiveness of gamification. Domínguez et al. (2013, p. 391) are, on the whole, positive about the utility of gamification but point out that “For many, the system was not motivating. In some cases the system was even discouraging.” They call for further research to identify the factors that cause this dissatisfaction.

Before we rush to incorporate gamified learning activities into the curriculum we need to understand its impact on engagement and learning from a student perspective. We also need to understand the contexts within which gamified learning activities work effectively. There is a clear need for a substantial body of further research to investigate these and other issues. Our study contributes to this agenda by conducting exploratory work which draws in the experiences of participants to identify empirically issues and challenges in gamified learning interventions.

**Research design**

In order to examine students’ perceptions of gamification, it was necessary to design and deploy a gamified learning activity. The activity used in this study is based on a prediction market (PM). A PM is “designed and run for the primary purpose of mining and aggregating information scattered among traders and subsequently using this information in the form of market values in order to make predictions about specific future events” (Tziralis & Tatsiopoulos, 2012, p. 75). In its simplest form, a contract is created whose value depends on a future uncertain event. For example, a manager may wish to evaluate whether a project will be completed on time. A contract is created which returns €100 if the project is completed on time and €0 otherwise. It is offered for sale at €50 (based on an initial 50:50 probability), typically on an electronic market. If market participants believe the project is likely to be completed on time, they will buy the contract causing its price to rise. If they believe the contrary, they will sell the contract, reducing the price. The price of the contract can therefore be used as an estimate of the group’s collective assessment as to the probability of the project being completed on schedule. The literature has previously established the utility of PMs as pedagogical tools that can be used across a range of disciplines (e.g., Buckley, Garvey, & McGrath, 2011; Buckley & Doyle, 2015; Evans, 2012).

This study was based on two modules focusing on the development of technical skills in calculating tax liabilities and students’ general knowledge of the national taxation system. One module is taken by a large third year undergraduate class (n = 142) taking a four year business degree, while the other is part of a one year full time graduate degree taken by a small group (n = 19) specialising in taxation. In addition to the obvious variance in terms of difficulty level, other differences between the cohorts which bear on the later analysis should be highlighted. Some undergraduate students pay an annual registration fee of €2,500 but otherwise receive free education; many more are covered by various grant schemes and pay nothing. The postgraduate students paid €7,250 for their programme. The undergraduates undertake paid work experience for 9 months directly after the semester in which the module is taken (called “co-op”). One of the key features of the postgraduate programme is that it offers numerous exemptions from the professional exams of the relevant national tax institute, assuming students attain sufficiently high grades.

In addition to developing technical skills in calculating tax liabilities, a learning outcome of both modules is that students develop their knowledge of how national taxation policy is developed and implemented. In order to meet this learning outcome, a tax PM was developed and deployed in both modules, with 10% of a student’s final grade being determined by their participation and performance. The Minister for Finance annually announces a range of tax policy decisions as part of the National Budget. The National Budget Forecasting Project (NBFP) required students to forecast what measures would be introduced in the forthcoming budget.
In the outcome they ying of the market. Over that period, students made an average of 48 trades. Each guide was developed. This followed a semi to act as a moderating limited set of and, PMs provide rapid feedback. At any point, the market price primary concerns with focus group methodology centre firstly on the environmental and NGO reports and position papers and recommendations from consultancy firms. Reading and analysing these should improve students’ general knowledge of tax, tying the activity back to the learning outcome. Grant Thornton sponsors the NBFP by paying for the relevant software license and by offering prizes from €500 to €100 for the top student performers, adding financial motivation. 161 students traded on the NBFP market. Students were free to trade at any stage throughout the three week period of the market. Over that period, students made an average of 48 trades. Each trade represents a student making or adjusting a forecast.

The PM encompasses all the elements of gamification. Implicit in the concept of market driven forecasting systems is the concept of an individualised reward. First, when participants correctly forecast future events they receive virtual cash increasing their portfolio value and their project grade. There was also a financial reward associated with the top performances. Second, PMs provide rapid feedback. At any point, the market price represents the consensus of all participants as to the probability of the relevant event occurring. A participant can therefore compare his/her personal estimates to the estimates of the entire class. Unlike a poll, a participant is not limited to making a single estimate. He/she can change decisions at any time in response to feedback or newly revealed information by buying or selling contracts. Third, at an operational level, students have a limited set of options. Contracts can be bought or sold. The complexity of the system arises from information aggregation and the repeated interactions of large numbers of traders. Finally, PMs are competitive; individuals can be ranked in terms of performance by comparing their portfolio values.

Research method

The aim of this study was to explore students’ perceptions of the gamified learning intervention outlined above. The most appropriate methodology was therefore to conduct focus groups. Focus groups originated in the 1920s in the area of market research but are currently a popular method of data collection in many fields (Robson, 2002). A focus group involves a group interview on a specific topic – namely, the “focus.” The idea is that people known to have had a certain experience can be interviewed in a relatively unstructured way about that experience (Bryman & Bell, 2003). Typically these open-ended group discussions are facilitated by a moderator and take at least an hour (Robson, 2002). The optimum group size is usually considered to be somewhere in the range from 8 to 12 participants (Stewart & Shamdasani, 1990).

Because of the number of participants involved in a focus group, a wide variety of viewpoints and ideas may emerge on any one issue, thus helping to explain or explore concepts (Saunders, Lewis & Thornhill, 2009). Individuals contributing to the discussion will often argue with each other and challenge each other’s views. This process causes more realistic accounts of people’s thoughts to emerge, because they are forced to think about and possibly revise their views (Bryman & Bell, 2003). Furthermore, participants tend to act as a moderating influence on each other so that extreme views are weeded out (Robson, 2002). The interaction of the group thus leads to a highly productive discussion which leads to a rich flow of data (Saunders, Lewis & Thornhill, 2009). The focus group approach therefore facilitates an observation of the ways in which individuals collectively make sense of a phenomenon and construct meanings around it (Bryman & Bell, 2003), making it an ideal research method for this exploratory study.

As participants in the gamified learning intervention, students’ expert knowledge, experience and perceptions are essential to advancing our understanding of the pedagogical impact of gamified interventions. Following research ethics approval, a focus group topic guide was developed. This followed a semi-structured, open ended format, and was developed using guidelines suggested by Wilkinson (2008). A series of open ended questions were developed to prompt discussion. The primary concerns with focus group methodology centre firstly on the...
ability of the moderator to manage and facilitate the group discussion, including encouraging involvement from all participants, and secondly, on the manner in which the data is recorded (Robson, 2002). In order to address these concerns an external moderator, highly experienced in focus group methodology, was recruited to conduct the focus groups. Independence meant the moderator was, and was perceived by the students to be, completely unbiased and unconnected with the learning intervention.

After the teaching semester concluded, all students in both relevant modules were sent an email inviting them to participate in a focus group discussion. Involvement was incentivised by offering participants university book store vouchers and refreshments. All students who volunteered to be involved were accepted into the study. Two separate focus groups were held, one for each of the two cohorts described in the research design section. The first group (F1) consisted of 13 business undergraduates (total class = 142). Six of this group were female and seven were male with the age range being between 20 and 24. The second group (F2) consisted of 9 postgraduate students (total class = 19) - five females and four males aged between 21 and 38 (one mature student represented a slight outlier in terms of age with the other 8 students being under 27). Each focus group lasted for just under 90 minutes. In line with best practice, both focus groups were recorded to ensure all data was accurately captured. During focus groups, students were engaged in a funnelled conversation beginning with general views of learning, moving to their perceptions of the PM, and finally involving in-depth discussions around the idea of learning through a gamified platform.

After data collection, the audiotapes were converted into verbatim transcripts before coding and analysis. The transcripts were read and reread to develop a full understanding of the discussions. The analysis focussed on identifying the key themes emerging. Unique topics were assigned a code, and analysis continued until no new categories emerged. Patterns and commonalities among categories were then identified and grouped into higher order themes. Once this was completed, transcripts were interrogated again with reference to the identified themes.

**Research findings**

Six themes emerged: learning outcomes, motivation, perceived stakes, group dynamics and gender, and challenges. Each is discussed below.

**Learning outcomes**

The undergraduate group displayed a greater understanding of the key aims of the activity and the lessons from it. First, they took the central message from the exercise to be a wider understanding of how the real world (as opposed to theory) operates, rather than anything tax specific, though some also acknowledged that their general tax knowledge had been enhanced.

> It might make you understand the reasoning why certain things are done in the budget. Before this, I used to look at it and go “why did they bring that in? It doesn’t make any sense.” But when you have to do the research behind it, you’re going “OK, that’s obviously going for a certain element that they’re changing for this reason.” UG

The postgraduate group displayed a more myopic attitude towards the project. They felt it was irrelevant for future tax professionals and better suited those aspiring to a career in trading. They failed to see a link between the assignment and the learning outcomes, and instead focused on the mechanics of trading. When the purpose of the assignment was pointed out by the moderator, they grudgingly saw the relevance; however they had a negative emotional response to the project.

> Well I personally think it had no use being in the Masters of Taxation, because working in practice, we’re never going to be on stocks, making trades or buying stocks. PG

Most students felt the project altered their information consumption habits. The budget is covered extensively in the media in the run up to budget day. Some students actively sought out new information sources.

> Well you looked for specific information relating to the question. PG
> …everyday it was a specific thing, like on VAT or income tax bands or something, you had to go tearing off to find out that area and come back with the answer and make a call on it. UG
Others reported that the project caused them to attend in a more focused way to information sources they would usually encounter. The undergraduates reported reading more than they would have otherwise.

*I looked up loads that I wouldn’t have been bothered with before.* UG

The postgraduates felt they did not read any more than they would have, mostly due to being at absolute capacity in terms of workload.

*You didn’t have time to be setting aside... because you’d so much other things going on, that you couldn’t let your other stuff suffer because of it.* PG

Participants identified two other key learning outcomes. First, they felt the project emphasised that real-world problems may not have predictable answers. A PM requires students to forecast inherently uncertain future events. The undergraduates recognised and appreciated this uncertainty. The postgraduates recognised that there was no right answer to the exercise, but were frustrated rather than intrigued by this.

The undergraduates also felt that the project illustrated herd mentality and group dynamics. Participants learned that the market could be influenced by the predictions of early movers. They appreciated that, crudely, it was a popularity contest for opinions. The undergraduates felt the market mechanism illustrated how group consensus can be manipulated and/or wrong, a valuable lesson in a business context. They appreciated this.

*...if you wanted the market to move in a certain way, all you’d to do was Google and find the thing that would support your argument and then the sheep effect would just kick in.* UG

The postgraduate group did not display such learning.

**Motivation**

The second theme was the impact the gamified intervention had on motivation. Here again there was a clear divide between groups, with the undergraduates being more favourably disposed towards the competitive element.

Students who were positive about competition commented favourably on a number of items. They liked the ranking system. They enjoyed seeing themselves climbing past “the competition” as the project progressed. As the ranking was visible to all, there was an extrinsic competitive motivation, where achievement was experienced because of relative positioning.

*I was monitoring my position the whole way along and that’s what was motivating me.* UG

Another motivation was the internal satisfaction of being able to “beat the system.” Students enjoyed learning how the market worked and could be influenced. Some attempted to manipulate others by posting misleading comments when they rationalised trades. They found the process of developing a strategy to allow them outsmart others rewarding. These students engaged in market manipulation in order to “win” but enjoyed the aspect of manipulation at an internal level.

*There was one question where I was the very first buyer on it and, you’re going to hate for me this, I bought a hundred shares in what I knew was going to be wrong, so people would see that spike in price. They all bought after me, sold them out and I probably screwed over about 20 undergrads.* PG

However, not all students were favourably disposed towards the competitive elements of the project. In particular, the majority of the postgraduates did not enjoy the competition.

One of the clear motivating factors for those who enjoyed the project was its novelty. For the undergraduates in particular, it represented a welcome break from the traditional learning activities they encountered.

*When a new question came out, everyone was rushing to their phones and stuff and you were kind of obsessed by it and you wanted to do really well.* UG
I think there was more of a “want” to do it, as opposed to a “need”… There was definitely an element of fun to it. UG

Many of the undergraduates were also motivated by the project’s real-time nature. They made a distinction between “time” and “timing.” Many took the view that it was not time consuming but about working smarter not harder. The undergraduates considered the project “always on,” but not a significant drain on time.

…the fact you could do it any time like, it was on your phone, so if you were just for instance on a bus or anything, you could just check it up… UG

The postgraduates agreed with the novelty of the approach. However, they found it frustrating to shift learning modalities during what they considered a very full schedule. They preferred to manage their time in blocks, planning in advance, and the PM did not fit this model.

There’s too much time gone into it for the 10%. PG

Many students were motivated by the financial prizes. They are accustomed to college work being something that has to be done to get good grades. The notion that it might be financially rewarding was both novel and motivating.

…the main motivation was money. If the money was not there I don’t think I would have spent so much time on it. PG

That said, the strength of the motivation was reduced by the calculated chance of success. Given the number of students involved, the statistical possibility of winning was under 10%. The greater the pool of possible winners, the less “real” the financial motivation was.

Perceived stakes for students

Consideration of the perceived stakes emerged as the single most important determinant of the suitability of gamification for a learning context. One of the major items highlighted repeatedly was the difference in the undergraduate and postgraduate perceptions of the project. The explanation emerging from discussions was the higher the stakes, the less open students were to a gamified approach.

The postgraduates look at their degree as a transaction. They pay approximately €7,250 and in return (as they see it) they are educated to a level where they earn significant professional exemptions and become more attractive entry level employees. Anything connected with exemptions gives rise to high stress levels. Many feel that without exemptions, the qualification is meaningless so stakes are high.

Well you’re after paying that much money, you’d want to get your exemptions, because if not it’s really you’ve just wasted €7,000 on this course. PG

They want the best possible chance of maximising marks in exemption driven modules. As such they are looking for three things in assessments; direct relevance to their future career, ease of completion and a way to maximise marks. The uncertainty associated with the PM does not represent the best opportunity to maximise marks. As the stakes were higher for them (in their view), they considered it unfair that they were assessed on a volatile “game.”

So say you lost a load of money in shares that you bought, so that means you could have lost an entire 5%, so I think that was an awful lot to lose in an exemptions module. PG

The undergraduates did not share that view. They were studying a general degree likely to contain elements they have no interest in. They could, therefore, appreciate how interesting subject matter can be in a gamified mode. While the purpose of college for them is to secure employment, this is not as specific, imminent or transactional. They are learning wider academic and life lessons and are less pigeon-holed, giving them a less focused view. As a result, they are under less pressure than the postgraduates in three ways. First, they have less intense course work (in both their viewpoint and that of the postgraduates). This prevented them from seeing the PM as being a disproportionate level of work. Second, they had not framed their degree programme in a transactional manner. They did not have the feelings of entitlement that the postgraduates did because they had not paid as much
money. Third, they were not as close to imminent employment and their proximity to co-op placement allowed them to feel there are strong prospects of eventual employment. In short, the fact that they were at a more general level of study with less at stake allowed them to gain more from the project.

**Group dynamics and gender**

Group dynamics were important. Discussions suggested that gamification worked better in larger, more anonymous groups where a close dynamic is not upset by competition. Some of the problems the postgraduates had stemmed from this. As a small class they typically studied and socialised together. Their closeness made them uncomfortable with the explicit competitive element of the project. The undergraduate class was larger, and lacked a common identity. This environment reduced the cognitive dissonance of competing with peers.

Gender differences also emerged as an issue. There was a definite gender difference observed in the focus groups.

*Well men are more competitive...they definitely prefer gambling, yeah. PG*

While the sample was not large enough to make accurate assumptions in this regard, it was observed that the motivating nature of the competitive element was more strongly emphasised by the male participants. This is an aspect of gamification worthy of further study.

**Challenges**

In order to inform future deployments, we investigated the challenges presented by the gamified activity. There was a sense from all participants that gamified interventions would not suit traditionally bookish students, due to the lack of “correct” answers and the proximity to the real world.

*Well I hang out with two girls and they were annoyed all the way through it...they are perfectionists and they could not get a handle on this because it came down to having fun, you know participating. They just couldn’t do it, they hated it. UG*

*It gives people who might not be good at like essays, you know, the typical learning, it gives them a chance to actually engage. PG*

A key challenge in designing competitive learning interventions is to ensure that motivation does not ebb because of competition. For students who did not begin well and were ranked towards the bottom early, motivation quickly decreased. In a gamified learning environment, students need to feel “in touch” with the leaders on the score-sheet in order to remain motivated.

*…if you knew you weren’t doing well earlier on, you kind of lost the motivation to stay going because you knew you weren’t going to get the right high marks. PG*

**Conclusions**

The study described in this paper had three objectives. Our first concern was to investigate students’ perceptions of the effect of gamification. Our research suggests that the effect is contextual. In a large undergraduate module taught as part of a general business degree, the gamified intervention engaged students because of its novelty and increased student motivation by introducing competition and rewards. This cohort learned how tax policy works, sought out new sources of information and employed strategy to enhance their performances. They also learned to appreciate the dangers of herd mentality and the fact that there isn’t always a correct answer in a real world context. Students in a small class undertaking a postgraduate module on a specialised programme did not achieve the same level of learning. They were frustrated by the gamified activity because they considered the stakes to be too high to engage in game playing, particularly in the context of a heavy postgraduate workload. This caused a negative response to the PM and a very myopic view of the learning outcomes it aimed to deliver. There does seem to be a perception that males were more engaged in the market, though it should be noted that the overall winner of €500 was female. The project also seemed to suit less bookish students better and most of the prize winners were not in the top group of academically performing students.
Our second goal was to provide guidelines for other practitioners in designing and implementing gamified learning interventions. Our study shows that the following factors need to be carefully considered:

- Class size (small, close-knit classes may not be as motivated by competition as large groups, however in large classes, not being within sight of the leaders may demotivate);
- The stakes involved (this might include whether students are undertaking an undergraduate or postgraduate programme, whether there are fees or professional exemptions involved, the workload of the students etc.);
- Whether the module is part of a general degree with broad learning outcomes or a specialist programme with specific learning outcomes (gamification may work better in introductory learning environments);
- What other teaching and learning approaches are being deployed to suit different learner types (gamification may suit some learners better than others so a broad range is needed);
- The nature and visibility of rewards (grades, leader boards, prizes) and understanding the potential motivation drop off points when students lose sight of the rewards;
- The key learning outcomes that the educator wants students to achieve (technical content, strategy, real world context, experiential learning, etc.).

Gamification is a novel concept in the higher education domain. While appearing to possess great potential as a pedagogical methodology, there are undoubtedly issues that should be studied in more detail before definitive conclusions can be drawn. Our third contribution is to begin to identify issues and questions that merit further study. Our research shows that context is a crucial determinant of the success of gamification. Factors such as class size, educational level and perceived stakes influence its effectiveness. Identifying the absolute effect of these factors, as well as other relevant contextual variables, would result in more optimal deployment of gamification in a higher education context. This research also suggests that gamified learning interventions suit some students and their learning styles better than others. Investigating and mapping this effect would facilitate gamification being used as part of a suite of learning interventions that deliver improved outcomes for all learners.

References


Exploring the Effects of Online Academic Help-Seeking and Flipped Learning on Improving Students’ Learning

Wen-Li Chyr¹, Pei-Di Shen², Yi-Chun Chiang³, Jau-Bi Lin⁴ and Chia-Wen Tsai⁵

¹Department of Information Management, Ming Chuan University, Taiwan // ²Institute of General Education, Ming Chuan University, Taiwan // ³icchien@mail.mcu.edu.tw // ⁴pdshen@mail.mcu.edu.tw // ⁵jblin@mail.mcu.edu.tw // jawen12b@gmail.com

(Submitted December 17, 2015; Revised May 5, 2016; Accepted May 9, 2016)

ABSTRACT
This study explored the effects of online academic help-seeking (OAHS) and flipped learning (FL) on students’ development of involvement, self-efficacy, and self-directed learning. A quasi-experiment was conducted to investigate whether students’ involvement, self-efficacy, and self-directed learning increases over time with intervention by OAHS, FL, and their combination. Three classes of first-year university students in a one-semester course were chosen for this empirical research. The 102 students were divided into three groups. The first group (G1, which received online OAHS and FL), and the second group (G2, which received online FL only), were the experimental groups. The last group (G3), which received the traditional teaching method in a blended course, served as the control group. The results indicate that G1 students’ involvement, self-efficacy, and self-directed learning all improved under the condition of simultaneously applying OAHS and FL. In addition, this study also reveals that the application of FL alone could be helpful in G2 students' development of their involvement, self-efficacy, and self-directed learning. However, G3 students, who learned with traditional teaching method in a blended learning environment, did not have better development in their involvement, self-efficacy, and self-directed learning. Finally, the authors further discuss the implications for teachers, scholars, and schools engaged in online education.

Keywords
Online academic help-seeking, Flipped learning, Involvement, Self-efficacy, Self-directed learning

Introduction

The needs for online academic help-seeking and flipped learning

Over the past two decades, teaching and learning processes have been influenced by technological, instructional and pedagogical advances (Chou & Tsai, 2002; Kavanoz, Yüksel & Özcan, 2015). Nowadays, students’ demands are transforming because their study habits and learning strategies have already changed due to the pervasiveness of the Internet (Persico & Pozzi, 2015). Regardless of which teaching or learning mode is used, students’ learning outcomes are one of the most important things for educators. In order to keep up with the rapid evolution of the education environment, teachers must update themselves on the potential of new teaching approaches frequently, and apply those to their instruction. Even though computers, digital tools, and educational technologies have been deemed as benefits to the education field, the potential advantages are not comprehensively understood (Wu, Kuo, Jen & Hsu, 2015).

As a result of the educational revolution, many colleges and universities now offer online programs or digital courses (Wei, Peng & Chou, 2015). Convinced that information technology (IT) can improve their teaching quality, relationships with their students, and provide students with effective educational experiences, many teachers have devoted themselves to apply IT and make effective use of it in class (Persico & Pozzi, 2015). A blended learning environment such as flipped classroom allows students to discover their own problems, encourages them in active learning and to have an open-minded attitude to create an atmosphere of cooperative learning (Tsai, Shen & Lu, 2015). Moreover, a web-based learning environment is helpful for improving learners’ help-seeking behaviours and influences their learning processes (Mäkitalo-Sieg! & Fischer, 2011). Therefore, students’ online learning experience and related processes is a popular research topic, and has recently been investigated by several teams (Roby, Ashe, Singh, & Clark, 2013; Tsai & Tsai, 2013). However, online learning also introduces some difficulties; for example, students may suffer alienation and isolation when they study in an online environment (McInnerney & Roberts, 2004; Tsai, 2013a). Thus, the authors of this study applied online academic help-seeking (OAHS), which refers to the spontaneous behavior of students requesting assistance from others or peers through the Internet (Cheng & Tsai, 2011), and investigated the effects on students’ learning in a flipped course.
In the 1990s, it was found that students felt physically isolated when they participated in online courses (Cereijo, Young & Wilhelm, 2001; Daugherty & Funke, 1998), especially when the instructor could not immediately provide feedback to learners (McIsaac, Blocher, Mahes, & Vrasidas, 1999). This problem remains till today; students suffer isolation when they study in an online environment and this situation is often considered to be unavoidable (McInerney & Roberts, 2004). In Taiwan, most students of compulsory education are taught by didactic, or spoon-fed, education. Upon entering college and participating in an online course without teacher’s on-the-spot support, students may not concentrate on learning materials, especially when seduced by potential distractions such as playing online games, surfing shopping websites, watching online series, and being addicted to social networks (Tsai, 2013b).

The importance of students’ involvement, self-efficacy, and self-directed learning

Online, blended learning, or flipped learning (FL) provides flexibility and accessibility. Different from traditional teaching approach, FL refers to instructors asking students to watch prescribed videos before class with other teaching materials to acquire knowledge and basic concepts, while the following in-class time is devoted to exercises, projects, or discussions of that content (Davies, Dean, & Ball, 2013). However, it is difficult for teachers to involve students in an online or blended course in an environment that is full of shopping websites, online games, and social networking websites (Tsai, 2012a). Flipping the classroom is a teaching approach that focuses on students’ learning involvement (McCallum, Schultz, Sellke, & Spartz, 2015). It is mentioned in earlier research that successful students’ involvement, which can be defined as a person’s perceived relevance of a certain object based on inherent needs, values, and interests (Zaichkowsky, 1985), could play a critical role in helping to elevate learning quality (Learning and Skills Council, 2007). Moreover, it is found that there are strong positive comments from students involved within a flipped class (McCallum et al., 2015). Therefore, the authors adopted FL in this study and measured whether students’ involvement is improved in a FL environment with OAHS.

From the educational perspective, students may need more self-efficacy experiences to enable them to learn successfully (Chen, Tutwiler, Metcalf, Kamarainen, Grotzer & Dede, 2016). Self-efficacy, which includes the belief that an individual has the ability to create change by personal action (Bandura, 2004), is also a critical factor in online education. It is indicated that positive Internet attitudes and preferences for web-based learning environments can be predicted by Internet self-efficacy (Chu & Tsai, 2009; Joo, Bong & Choi, 2000). Moreover, self-efficacy has been shown to influence students’ motivation and learning outcomes (Liang & Tsai, 2008; Tsai, 2012b). Furthermore, compared with traditional lecture-based learning, FL can facilitate students’ cognitive engagement and guide them to interact more efficiently with the learning content (Ibrahim & Callaway, 2014). That is, it is important to improve students’ self-efficacy in an online or blended course. In order to understand learners’ self-directed use of technology for learning, it is necessary to understand what self-directed learning entails (Lai, 2015). In an online learning environment, students have to access the course independently and structure the time, pace, and strategy of their own learning (Puzziferro, 2008). Thus, it is important to develop students’ self-directed learning to help them learn well at their own pace, anytime and anywhere (Tsai, Shen, & Huang, 2012). Knowles (1989) defined self-directed learning as “a process in which individuals take the initiative, with or without the help from others, in diagnosing their learning needs, formulating goals, identifying human and material resources, choosing and implementing appropriate learning strategies and evaluating learning outcomes” (p. 18). In flipped classrooms, learners feel more confident during course discussions because they have already previewed and prepared the learning materials before class; this can prompt their self-directed learning for their learning activities (Bishop & Verleger, 2013; Halili, Razak & Zainuddin, 2015).

In previous studies related to computing education, this research team emphasized the cultivation of students’ computing skills and the pass rate on computing certifications (Lee, Shen & Tsai, 2010; Shen, Lee & Tsai, 2011; Tsai, 2016; Tsai & Shen, 2014; Tsai, Shen & Tsai, 2011); these studies showed that it is also important to understand how to improve students’ learning psychology, such as their involvement and self-efficacy. Therefore, the authors designed and integrated OAHS in the implementation of FL in a blended course titled “Applied Information Technology: Office Software,” and explored their effects on improving students’ involvement, self-efficacy, and self-directed learning.
The current study

Subjects

The current study provides an analysis of three classes of first-year university students in Taiwan (27 male and 75 female), and taking a compulsory course titled “Applied Information Technology: Office Software.” Two classes were from the department of Finance, while one was from the department of Law. An experienced professor taught the three classes during the same semester, with each of the classes taught using a different teaching approach. In this study, G1 (which received treatments of online academic help-seeking and flipped learning, \( n = 33 \)) and G2 (which received treatment of flipped learning only, \( n = 34 \)) were the experimental groups, while G3 group (which received traditional teaching, \( n = 35 \)) served as the control group.

Course setting

The experiment mainly targeted first-year students from non-computer or non-information departments of study, and was conducted in a semester-long, two credit-hour course entitled “Applied Information Technology: Office Software.” Computing education is emphasized for students of each education level in Taiwan. Even students in university departments of Applied Japanese or Law are still required to take four compulsory computing courses before they graduate (Tsai, 2012a). Thus, this course aims to develop students’ computing skills for using document processing software such as Microsoft Word and PowerPoint.

Treatments in this study

Three classes, each treated as a group, were involved in the experiment. The first class (G1) received the treatment of online academic help-seeking and flipped learning simultaneously. The second class (G2) received the treatment of flipped learning only. The last class (G3) received the traditional teaching approach, but in a blended learning environment.

Treatment of OAHS

According to an official report, 84.9% of Taiwanese Internet users aged 12 and above have smartphone-enabled Internet access (The National Development Council, 2014). In order to make sure that this study could be conducted well, students in G1 were divided into small teams at the beginning of the semester, each of the groups consisting of four to five members. In addition, the teacher also investigated smartphone ownership rate in class, and found that every student in this study possessed a smartphone, and regarded a smartphone as their main device for accessing the Internet. Moreover, all of the students had a Facebook account and had also downloaded and installed a mobile application called “LINE.” Therefore, each group of students was required to form an online learning community through which group members could raise questions, discuss, share information and remind one another to submit homework.

In the process of help-seeking through LINE, the teacher was not directly involved, nor did he provide answers for students. Instead, the members of every team in G1 provided assistance for each other. Thus, the students who faced problems did not have to worry if they wasted teacher’s and classmates’ time in the classroom. In addition, students could post tips in the “Notes” section of LINE. They could even directly talk to their teammates via video conference in private or in the team group via LINE. Moreover, students had to take screenshots of online discussion, help-seeking, and problem-solving, then upload them to the course website every week. In that way, the researchers in this study could confirm that students really adopted OAHS and solved their problems.

Treatment of FL

In the implementation of FL learning approach, the authors adopted and followed Datig and Ruswick’s (2013) suggestions that students should preview the teaching materials outside of class so that they can prepare themselves well for discussion in class. In this study, students in G1 and G2 were asked to watch two or three
video lessons with flash, video, and voice (each lasted 10 to 20 minutes) before class every week to help them absorb basic knowledge. Students could use computer or smartphone for this task.

In class, students were asked to discuss, share, reflect on, and practice what they learned from the teaching videos. In order to understand the effectiveness of learning, the teacher arranged short quizzes during the class, and regarded this as formative assessment. The teacher in this study raised questions for students’ discussion and interaction in the class time. The teacher also selected students for asking questions to confirm whether students watched the teaching videos. Moreover, the teacher kept track of students’ learning outcomes, solved problems that arose, and improved teaching strategy based on the quiz results.

**Measurements**

**Students’ learning performance**

In this computing course, the researchers measured students’ computing skills as their learning performance. In this study, students were required to take an examination for a certificate in Microsoft PowerPoint in the eleventh week and one for Word in the seventeenth week of the semester. These examinations were administered by Computer Skills Foundation in Taiwan. On the examination, students had forty minutes to complete the simulation problems. A student’s score comes from her/his correctness and completeness of problem solving. A surrogate sum representing computing skills was averaged from the scores of these two examinations. Then, the authors tested the skills differences among the three groups of students as the measure of their learning performance.

**Students’ involvement, self-efficacy, and self-directed learning**

In this study, the researchers used a quasi-experimental design and questionnaire to measure students’ involvement, self-efficacy and self-directed learning. In order to measure students’ involvement in the blended course, Zaichkowsky’s (1994) Personal Involvement Inventory (PII) was adapted, which comprised ten items rated by the students about themselves on a seven-point scale used to evaluate about their personal psychological states. Besides, the Motivated Strategies for Learning Questionnaire (MSLQ), composed of seven items self-rated on a five-point scale, was adopted to investigate student’s self-efficacy for learning and performance. Finally, the Self-directed Learning Readiness Scale (SLDRS) by Guglielmino (1977) was adopted for examining students’ self-directed learning ability. SLDRS is a 58-item, 5-point Likert-type scale, where a score of 1 denotes “strongly disagree” and a score of 5 denotes “strongly agree.” The higher scores students got, the better self-directed learning they exhibited.

All students were required to complete these three scales in the first week as a pretest, and then complete them again at the end of semester as a posttest. Students who did not complete the pretest, posttest and certificate examinations were removed from analysis in this study. At the end, 102 students, who completed all questionnaires and took the certificate examinations, were evaluated as participants in this study.

**Results**

**Pretest**

According to the analysis of pretests shown in Table 1, the difference of students’ involvement, self-efficacy and self-directed-learning among G1, G2, and G3 are not significant statistically. Moreover, the authors also checked students’ computing skills of using Microsoft Word or PowerPoint before the course began. In the first week of the semester, the teacher in this course asked if students had previously learned or used Microsoft Word or PowerPoint. The students who had learned Microsoft Word or PowerPoint were excluded from the experimental sample, although they still remained in this course. Based on the analysis in the pretest and teacher’s check, it is believed that the participating students had equal levels of computing skills, involvement, self-efficacy and self-directed learning when the experiment started. Therefore, the potential threat of initial variance among students’ computing skills, involvement, self-efficacy and self-directed learning can be excluded.
Table 1. One-way ANOVA: Pretest of students’ involvement, self-efficacy and self-directed learning

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Group (I)</th>
<th>Group (J)</th>
<th>Mean difference (I-J)</th>
<th>Std. error</th>
<th>Sig.</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involvement</td>
<td>G1</td>
<td>G2</td>
<td>.16391</td>
<td>.11674</td>
<td>.377</td>
<td>.914</td>
<td>.404</td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G2</td>
<td>.12072</td>
<td>.11585</td>
<td>.583</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>.00531</td>
<td>.03941</td>
<td>.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>.01272</td>
<td>.03911</td>
<td>.949</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G2</td>
<td>-.01593</td>
<td>.09123</td>
<td>.985</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G1</td>
<td>G2</td>
<td>.09815</td>
<td>.09053</td>
<td>.558</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>.00531</td>
<td>.03941</td>
<td>.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>.00742</td>
<td>.03969</td>
<td>.983</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G2</td>
<td>-.01593</td>
<td>.09123</td>
<td>.985</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>.11408</td>
<td>.09188</td>
<td>.465</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>G1</td>
<td>G2</td>
<td>-.16391</td>
<td>.11674</td>
<td>.377</td>
<td>1.069</td>
<td>.347</td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G2</td>
<td>-.04319</td>
<td>.11757</td>
<td>.935</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>-.00531</td>
<td>.03941</td>
<td>.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>.00742</td>
<td>.03969</td>
<td>.983</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G2</td>
<td>-.00742</td>
<td>.03969</td>
<td>.983</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>.01593</td>
<td>.09123</td>
<td>.985</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>-.01593</td>
<td>.09123</td>
<td>.985</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>.11408</td>
<td>.09188</td>
<td>.465</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-directed</td>
<td>G1</td>
<td>G2</td>
<td>-.12072</td>
<td>.11585</td>
<td>.583</td>
<td>.053</td>
<td>.948</td>
</tr>
<tr>
<td>learning</td>
<td>G3</td>
<td>G2</td>
<td>.04319</td>
<td>.11757</td>
<td>.935</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>-.01272</td>
<td>.03911</td>
<td>.949</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>-.00742</td>
<td>.03969</td>
<td>.983</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G2</td>
<td>-.00742</td>
<td>.03969</td>
<td>.983</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>.11408</td>
<td>.09188</td>
<td>.465</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Posttest

In this study, the Paired-samples t-test was applied to compare students’ involvement in this blended computing course in G1 (OAHS and FL), G2 (FL), and G3 (traditional teaching method). As the results show in Table 2 and Table 3, it is revealed that the group which received OAHS with FL (G1) showed a significant increase in students’ involvement at the end of the semester (mean = 4.4333) in contrast to their pretest involvement (mean = 4.1788) (p < .05). Moreover, G2, who received FL alone, also had a significant increase in students’ involvement by the end of the semester (mean = 4.3059) in contrast to their pretest involvement (mean = 4.0647) (p < .05). These results suggest that OAHS and FL could have a positive effect on students’ involvement. As for G3, who received the traditional teaching approach, the data show no significant difference between pretest (mean = 4.1629) and posttest (mean = 4.1057) (p = .637).

Table 2. Paired samples statistics: Involvement

<table>
<thead>
<tr>
<th>Involvement</th>
<th>Pre-post</th>
<th>M</th>
<th>n</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>Pretest</td>
<td>4.1788</td>
<td>33</td>
<td>.34978</td>
<td>.06089</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>4.4333</td>
<td>33</td>
<td>.49666</td>
<td>.08646</td>
</tr>
<tr>
<td>G2</td>
<td>Pretest</td>
<td>4.0647</td>
<td>34</td>
<td>.32369</td>
<td>.05551</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>4.3059</td>
<td>34</td>
<td>.48738</td>
<td>.08359</td>
</tr>
<tr>
<td>G3</td>
<td>Pretest</td>
<td>4.1629</td>
<td>35</td>
<td>.44131</td>
<td>.07460</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>4.1057</td>
<td>35</td>
<td>.49464</td>
<td>.08361</td>
</tr>
</tbody>
</table>

Table 3. Pair-wise comparison of students’ involvement

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>t-value</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>.25455</td>
<td>.59007</td>
<td>.10272</td>
<td>2.478</td>
<td>32</td>
<td>.019</td>
</tr>
<tr>
<td>G2</td>
<td>.24118</td>
<td>.57951</td>
<td>.09938</td>
<td>2.427</td>
<td>33</td>
<td>.021</td>
</tr>
<tr>
<td>G3</td>
<td>-.05714</td>
<td>.71096</td>
<td>.12017</td>
<td>-.476</td>
<td>34</td>
<td>.637</td>
</tr>
</tbody>
</table>

Note. *p < .05.

In terms of self-efficacy, Table 4 and Table 5 show the average scores for both G1 and G2 illustrate a statistically significant increase on the posttest. G1 showed a significant increase in students’ self-efficacy at the end of the semester (mean = 4.3603) in contrast to their pretest self-efficacy (mean = 3.7915) (p < .001). Moreover, G2 also had a significant increase in students’ self-efficacy at the end of the semester (mean = 4.3947) in contrast to their pretest self-efficacy (mean = 3.8347) (p < .001). However, G3 showed no significant differences between pretest (mean = 3.9554) and posttest (mean = 3.7103) (p = .083).
Table 4. Paired samples statistics: Self-efficacy

<table>
<thead>
<tr>
<th>Self-efficacy</th>
<th>Pre-post</th>
<th>M</th>
<th>n</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>Pretest</td>
<td>3.7915</td>
<td>33</td>
<td>.53125</td>
<td>.09248</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>4.3603</td>
<td>33</td>
<td>.69644</td>
<td>.12123</td>
</tr>
<tr>
<td>G2</td>
<td>Pretest</td>
<td>3.8347</td>
<td>34</td>
<td>.50499</td>
<td>.08661</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>4.3947</td>
<td>34</td>
<td>.56165</td>
<td>.09632</td>
</tr>
<tr>
<td>G3</td>
<td>Pretest</td>
<td>3.9554</td>
<td>35</td>
<td>.40105</td>
<td>.06779</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
<td>3.7103</td>
<td>35</td>
<td>.66920</td>
<td>.11312</td>
</tr>
</tbody>
</table>

Table 5. Pair-wise comparison of students' self-efficacy

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>t-value</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>.56879</td>
<td>.76612</td>
<td>.13336</td>
<td>4.265</td>
<td>32</td>
<td>.000***</td>
</tr>
<tr>
<td>G2</td>
<td>.56000</td>
<td>.75792</td>
<td>.12998</td>
<td>4.308</td>
<td>33</td>
<td>.000***</td>
</tr>
<tr>
<td>G3</td>
<td>-.24514</td>
<td>.81310</td>
<td>.13744</td>
<td>-1.784</td>
<td>34</td>
<td>.083</td>
</tr>
</tbody>
</table>

Note. ***p < .001.

With regard to students' self-directed learning, it is shown in Table 6 and Table 7 that students in G1 had a significant difference between the scores for pretest (mean = 3.1521) and posttest (mean = 3.2679) (p < .05). These results suggest that OAHS and FL had a positive effect on developing students' self-directed learning. In addition, students from G2 also had significant difference between the pretest (mean = 3.1447) and posttest (mean = 3.2976) scores (p < .01). However, the control group (G3) did not have statistically significant difference in self-directed learning between their pretest (mean = 3.1574) and posttest (mean = 3.1820) (p = .590).

Table 6. Paired samples statistics: Self-directed learning

<table>
<thead>
<tr>
<th>Self-directed learning</th>
<th>Pre-post</th>
<th>M</th>
<th>n</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>Pretest</td>
<td>3.1521</td>
<td>33</td>
<td>.15562</td>
<td>.02709</td>
</tr>
<tr>
<td>G2</td>
<td>Posttest</td>
<td>3.2679</td>
<td>33</td>
<td>.13090</td>
<td>.02279</td>
</tr>
<tr>
<td>G3</td>
<td>Pretest</td>
<td>3.1447</td>
<td>34</td>
<td>.17664</td>
<td>.03029</td>
</tr>
<tr>
<td>G4</td>
<td>Posttest</td>
<td>3.2976</td>
<td>34</td>
<td>.25352</td>
<td>.04348</td>
</tr>
<tr>
<td>G5</td>
<td>Pretest</td>
<td>3.1574</td>
<td>35</td>
<td>.15413</td>
<td>.02605</td>
</tr>
<tr>
<td>G6</td>
<td>Posttest</td>
<td>3.1820</td>
<td>35</td>
<td>.22112</td>
<td>.03738</td>
</tr>
</tbody>
</table>

Table 7. Pair-wise comparison of students' self-directed learning

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>t-value</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>.11576</td>
<td>.24200</td>
<td>.04213</td>
<td>2.748</td>
<td>32</td>
<td>.010'</td>
</tr>
<tr>
<td>G2</td>
<td>.15294</td>
<td>.28911</td>
<td>.04958</td>
<td>3.085</td>
<td>33</td>
<td>.004**</td>
</tr>
<tr>
<td>G3</td>
<td>.02457</td>
<td>.26746</td>
<td>.04521</td>
<td>.543</td>
<td>34</td>
<td>.590</td>
</tr>
</tbody>
</table>

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

In order to understand students' learning performance after participating in this experiment, descriptive statistics were used. It is found that students in G1, who received the intervention of FL and OAHS, had the highest average score in computing skills (mean = 89.03) among the three groups. Following were G2, who received FL only, with the second highest average score (mean = 85.68). Though students from the control group (G3) received the lowest score in computing skill (mean = 84.29) among the three groups (see Table 8), the difference of students' computing skills among the three groups was not significant.

Table 8. Students' average scores

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>33</td>
<td>89.03</td>
<td>13.153</td>
<td>2.290</td>
</tr>
<tr>
<td>G2</td>
<td>34</td>
<td>85.68</td>
<td>13.847</td>
<td>2.487</td>
</tr>
<tr>
<td>G3</td>
<td>35</td>
<td>84.29</td>
<td>16.787</td>
<td>2.838</td>
</tr>
</tbody>
</table>

As illustrated in the posttest data, shown in Table 9, the differences of students' involvement and self-efficacy among the three groups were significant statistically. At the end of the semester, G1 students' involvement was higher than G3 in a statistically significant manner (p < .05). That is, students who simultaneously adopted OAHS and FL had better development of involvement that those did not. In addition, it is also found that students from the control group (G3), who received traditional teaching method, had the lowest degree of self-efficacy among the three groups (p < .05). However, the differences of students' self-directed-learning and their computing skills among G1, G2, and G3 are not significant statistically.
Table 9. One-way ANOVA: Posttest of students’ involvement, self-efficacy, self-directed learning, and computing skills

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Group (I)</th>
<th>Group (J)</th>
<th>Mean difference (I-J)</th>
<th>Std. error</th>
<th>Sig.</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involvement</td>
<td>G1</td>
<td>G2</td>
<td>.12745</td>
<td>.12045</td>
<td>.573</td>
<td>3.831</td>
<td>.025*</td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td></td>
<td>.32762*</td>
<td>.11959</td>
<td>.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>-.12745</td>
<td>.12045</td>
<td>.573</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td></td>
<td>.20017</td>
<td>.11869</td>
<td>.246</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>-.32762*</td>
<td>.11959</td>
<td>.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td></td>
<td>-.20017</td>
<td>.11869</td>
<td>.246</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>G1</td>
<td>G2</td>
<td>-.03440</td>
<td>.15756</td>
<td>.976</td>
<td>12.343</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td></td>
<td>.65002*</td>
<td>.15645</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>.03440</td>
<td>.15756</td>
<td>.976</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td></td>
<td>.68442*</td>
<td>.15526</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>-.65002*</td>
<td>.15645</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td></td>
<td>-.68442*</td>
<td>.15526</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-directed</td>
<td>G1</td>
<td>G2</td>
<td>-.02977</td>
<td>.05112</td>
<td>.844</td>
<td>2.849</td>
<td>.063</td>
</tr>
<tr>
<td>Learning</td>
<td>G3</td>
<td></td>
<td>.08588</td>
<td>.05075</td>
<td>.244</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>.02977</td>
<td>.05112</td>
<td>.844</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td></td>
<td>.11565</td>
<td>.05037</td>
<td>.077</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>-.08588</td>
<td>.05075</td>
<td>.244</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td></td>
<td>-.11565</td>
<td>.05037</td>
<td>.077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computing</td>
<td>G1</td>
<td>G2</td>
<td>3.353</td>
<td>3.688</td>
<td>.663</td>
<td>.920</td>
<td>.402</td>
</tr>
<tr>
<td>Skills</td>
<td>G3</td>
<td></td>
<td>4.745</td>
<td>3.578</td>
<td>.418</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>G1</td>
<td>-3.353</td>
<td>3.688</td>
<td>.663</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td></td>
<td>1.392</td>
<td>3.636</td>
<td>.929</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>G1</td>
<td>-4.745</td>
<td>3.578</td>
<td>.418</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td></td>
<td>-1.392</td>
<td>3.636</td>
<td>.929</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Discussion and implications

Computers and the Internet are regarded as valuable tools that provide a wealth of information and useful online resources for assisting with learning activity (Cheng & Tsai, 2011). As not all online classes are created equal, to understand how different modes of information are related to student experiences and learning in online classes, teachers can clarify what constitutes effective instructional design in online contexts (Clark & Mayer, 2008; Limperos, Buckner, Kaufmann & Frisby, 2015). Therefore, the present study adopted online FL and OAHS, and verified their effects on enhancing students’ involvement, self-efficacy and self-directed learning in a blended computing course.

According to the results, the authors believe this study could make some positive contributions to online learning environments in three different ways. Above all, this study reveals the effect of OAHS and FL, and how an instructor improved students’ involvement, self-efficacy, and self-directed learning under simulated conditions by applying OAHS and FL. Secondly, this study also indicates that the sole application of FL could be helpful in students’ development of their involvement, self-efficacy, and self-directed learning. Finally, this study may be one of the first attempts to simultaneously apply OAHS and FL in an online computing course to improve students’ learning.

The combined effect of OAHS and FL

The online learning environment not only changes teaching and learning strategies, but also provides new possibilities for academic help-seeking (Cheng, Liang & Tsai, 2013). With respect to the combined effects of OAHS and FL, the authors in this empirical study find support from the analyses presented in Table 3, Table 5, and Table 7. That is, there is a significant increase at the end of the course in G1 students’ involvement (\(p = .019\)), self-efficacy (\(p = .000***\)), and self-directed learning (\(p = .010\)). In addition, Table 9 shows that the effect of OAHS and FL results in a significant difference in students’ involvement (\(F = 3.831, p = .025 < .05\)) and self-efficacy (\(F = 12.343, p = .000 < .001\)). G1 students’ involvement and self-efficacy were better than
those in G3. According to the results in this study, it is believed that OAHS and FL could lead to better development of students’ involvement, self-efficacy, and self-directed learning.

The potential reasons for these results may be due to the Internet providing abundant information resources and flexible time and space for learning (Cheng, Liang & Tsai, 2013; Rouet, 2006). Students who are convinced that online information is an essential resource related to course content are willing to seek online academic help (Lee, Chiu, Liang & Tsai, 2014). Moreover, academic help-seeking behavior can be regarded as a kind of self-regulation learning strategy or self-efficacy, thus the high level of self-efficacy can affect the strategies of self-regulated learning (Cheng et al., 2011). In this study, it was found that even in a learning environment without teachers’ on-site monitoring, students’ online learning behavior can be effective with adoption of OAHS in a course. Therefore, it is suggested that teachers could combine and adopt OAHS and FL, so as to provide the opportunity for students’ to seek academic help online by themselves, obtain prior knowledge before class and further contribute to their involvement, self-efficacy and self-directed learning in a blended computing course.

The effect of FL

In this empirical research, the use of FL is found to play a positive role in developing students’ involvement, self-efficacy and self-directed learning in a blended computing course. As the data shows in Table 3, Table 5, and Table 7, there are significant differences by the end of the course in G2 students’ involvement ($p = .021^{*}$), self-efficacy ($p = .000^{***}$), and self-directed learning ($p = .004^{**}$) in the flipped course. In addition, data presented in Table 9 shows that G2 students’ self-efficacy was significantly higher than that of those in G3.

This study demonstrates that FL instruction is helpful and could result in a better development of students’ involvement, self-efficacy and self-directed learning.

In previous research, it is reported that peers play an important role in the FL environment, as they help to improve the learning process and to internalize knowledge, and that the levels of comprehension and capability to resolve problems are much higher in FL than for those who are taught by traditional teaching approach (Dasgupta & Tuttle, 2013; Crouch & Mazur, 2001). Results of the present study are similar to those of Sams and Bergmann’s (2013) and Tsai, Shen and Lu’s (2015) studies, which indicate that the flipped classroom could help improve students’ learning. Therefore, it is suggested that teachers could adopt FL to provide more channels for students’ learning and reflection, enhance their interaction, develop regular learning habits, and further improve students’ learning performance.

The effect of traditional teaching approach

Based on the data shown in Table 3, Table 5 and Table 7, there is no significant difference in G3 students’ involvement ($p = .637$), self-efficacy ($p = .083$) or self-directed learning ($p = .590$) by the end of the semester. Traditional lecturing is one-way knowledge transfer, and students are expected to memorize and comprehend the learning concepts (Herreid & Schiller, 2013; Tsai, Shen & Lu, 2015). Thus, without redesigning the course and teaching approach for students, traditional teaching, even in a blended learning environment, may not be an effective approach (Tsai, 2015).

Finally, according to the data shown in Table 9, there is a warning signal for teachers who plan to provide flipped courses and researchers who investigate the effects of online education. Teachers who wish to directly transform their traditional teaching methods into a digital presentation, without re-designing the course and teaching methods, may find it is hard to achieve satisfactory learning effects. For example, students who adopted neither OAHS nor FL, in G3, had the significantly lowest self-efficacy among the three groups (see Table 9). G3 students’ involvement, self-directed, and computing skills are lower than G1 and G2, though insignificantly. Based on our findings in this study, it is once again suggested that teachers should adopt innovative teaching approaches and redesign a course before directly providing a flipped or blended course for students, in order to fully exploit the benefits of the Internet and educational technologies (Tsai & Shen, 2009).

Limitations of this study

Although it is found that OAHS and FL improved students’ involvement, self-efficacy, and self-directed learning in this blended course, some other potential factors may cause bias when evaluating students’ learning
performance. For example, the potential effects of students’ readiness for online learning may also influence their performance (Tsai, 2012a). Despite the researchers conducting a pretest to check students’ involvement, self-efficacy, and self-directed learning when they first participated in this course, students’ characteristics in the three groups are not necessarily all the same. In addition, there may exist the possibility that the students did not engage voluntarily but were forced to follow the course protocol for increasing their involvement, self-efficacy, and self-directed learning. Therefore, the existing differences may influence students’ acceptance of blended course or traditional course and lead to bias of measurement (Shen, Lee, & Tsai, 2007).

Conclusion

With the appearance of technology-based learning, help seeking with information and communication technology tools has now become an important issue in contemporary learning environments (Lee, Chiu, Liang & Tsai, 2014). Students can search for online academic help-seeking through the plentiful resources on the Internet (Cheng, Liang & Tsai, 2013). In addition, the advantages of mobile learning and applications not only facilitate users to study anytime, anywhere, but also to get feedback immediately. Thus, how to use these benefits well and design an appropriate teaching approach is important (Huang, Hwang & Chang, 2010). It is indicated that many students may not log in their course website every day; however, they usually log in and browse social network sites or chat via mobile application every day (Tsai, in press). Thus, the authors utilized one of the most locally popular communication applications, LINE, and integrated FL and OAHS into an online computing course. Moreover, the researchers in this study measured the effects of innovative adoption of mobile technology and help-seeking on improving students’ learning psychology, such as their involvement, self-efficacy, and self-directed learning, in this online computing course. Therefore, the integration and implementation of OAHS and FL could provide comprehensive implications for educators to design their future online or blended courses and help their students to involve themselves in the course.

As FL and blended learning become trends in innovative teaching approaches, OAHS and FL have become important factors in contemporary learning environments. Therefore, in this present study, the authors adopted OAHS and FL to involve students in a blended computing course. The findings of this study reveal that students with interventions of OAHS and FL, and FL alone had significant increase in the development of their involvement, self-efficacy, and self-directed-learning. On the contrary, under a traditional teaching approach, students may not be well involved, nor have a high level of self-efficacy or self-directed learning in a blended computing course. The authors hope that these results regarding the implementation of OAHS and FL can provide comprehensive insights for instructors to assist students construct high levels of involvement, self-efficacy and self-directed learning in blended learning environments, particularly for computing courses.

References


Appendix A

Self-directed learning readiness scale

Items

1. I’m looking forward to learning as long as I’m living.
2. I know what I want to learn.
3. When I see something that I don’t understand, I stay away from it.
4. If there is something I want to learn, I can figure out a way to learn it.
5. I love learn.
6. It takes me a while to get started on new projects.
7. In a classroom, I expect the teacher to tell all class members exactly what to do at all times.
8. I believe that thinking about who you are, where you are, and where you are going should be a major part of every person’s education.
9. I don’t work very well on my own.
10. If I discover a need for information that I don’t have, I know where to go to get it.
11. Even if I have a great idea, I can’t seem to develop a plan for making it work.
12. In a learning experience, I prefer to take part in deciding what will be learn and how.
13. Difficult study doesn’t bother me if I’m interested in something.
14. No one but me is truly responsible for what I learn.
15. I can tell whether I’m learning something well or not.
16. There are so many things I want to learn that I wish that there were more hours in a day.
17. If there is something I have decided to learn, I can find time for it, no matter how busy I am.
18. Understanding what I read is a problem for me.
19. I don’t learn, it’s not my fault.
20. I know when I need to learn more about something.
21. If I can understand something well enough to get a good grade on a test, it doesn’t bother me if I still have questions about it.
22. I think libraries are boring places.
23. The people I admire most always learning new things.
24. I can think of many different ways to learn about a new topic.
25. I try to relate what I am learning to my long-term goals.
26. I am capable of learning for myself almost anything I might need to know.
27. I really enjoy tracking down the answer to a question.
28. I don’t like dealing with questions where there is not one right answer.
29. I have a lot of curiosity about things.
30. I’ll glad when I’m finished learning.
31. I’m not as interested in learning as some other people seem to be.
32. I don’t have any problem with basic study skills.
33. I like to try new things, even if I’m not sure how they will turn out.
34. I don’t like it when people who really know what they’re doing point out mistake that I am making.
35. I’m good at thinking of unusual ways to do things.
36. I like to think about the future.
37. I’m better than most people are at trying to find out the things I need to know.
38. I think of problems as challenges, not stop signs.
39. I can make myself do what I think I should.
40. I am happy with the way I investigate problems.
41. I become a leader in group learning situations.
42. I enjoy discussing ideas.
43. I don’t like challenging learning situations.
44. I have a strong desire to learn new things.
45. The more I learn, the more exciting the world becomes.
46. Learning is fun.
47. It’s better to stick with the learning methods that we know will work instead of always trying new ideas.
48. I want to learn more so that I can keep growing as a person.
49. I am responsible for my learning – no one else is.
50. Learning how to learn is important to me.
51. I will never be too old to learn new things.
53. Constant learning is a bore.
54. Learning is a tool for life.
55. I learn several new things on my own each year.
56. Learning doesn’t make any difference in my life.
57. I am an effective learner in the classroom and on my own.
58. Learners are leaders.
Web-Based System for Adaptable Rubrics: Case Study on CAD Assessment

Pedro Company1, Manuel Contero2, Jeffrey Otey3, Jorge D. Camba4, María-Jesús Agost5 and David Pérez-López2

1Institute of New Imaging Technologies, Universitat Jaume I, Castellón, Spain // 23IB, Universitat Politècnica de València, Valencia, Spain // 3Zachry Department of Civil Engineering, Texas A&M University, TX, USA // 4Gerald D. Hines College of Architecture and Design, University of Houston, TX, USA // 5Department of Mechanical Engineering and Construction, Universitat Jaume I, Castellón, Spain // pcompany@uji.es // mcontero@upv.es // j-otey@tamu.edu // jdorribo@uh.edu // magost@uji.es // dapelo@i3b.upv.es

*Corresponding author

(Submitted December 23, 2015; Revised August 16, 2016; Accepted August 23, 2016)

ABSTRACT

This paper describes the implementation and testing of our concept of adaptable rubrics, defined as analytical rubrics that arrange assessment criteria at multiple levels that can be expanded on demand. Because of its adaptable nature, these rubrics cannot be implemented in paper formats, neither are they supported by current Learning Management Systems (LMS). The main contribution of this work involves the adaptable capability of different levels of detail, which can be expanded for each rubric criterion as needed. Our rubrics platform provides specialized and intuitive tools to create and modify rubrics as well as managing metadata to support learning analytics. As an example of a practical assessment situation, a case study on Mechanical Computer Aided Design (MCAD) systems training is presented. The validation process in this scenario proved the effectiveness of our adaptable rubric platform for supporting formative assessment in a multifaceted and complex field such as MCAD. The system also showed the potential of collecting user interaction metadata, which can be used to analyze the evaluation process and guide further improvements in the teaching strategy.

Keywords
Adaptable rubrics, Rubrics platform, Formative assessment, Learning Management Systems

Introduction

Rubrics are a recognized instrument to support authentic assessments to describe student achievement (Andrade, 1996; Andrade, 2000). A rubric can be defined as a scoring tool that provides a set of criteria to assess a piece of work and includes gradations of quality or performance for each criterion. Rubrics can increase student-learning outcomes by making teachers’ expectations explicit and by showing students how to meet these expectations (by presenting what level of quality is expected from their work). Rubrics are also useful to help students develop a critical sense of their own work by providing them with criteria to become more thoughtful judges of the quality of their own and others’ work.

True assessment emphasizes the application and use of knowledge to solve complex tasks that involve contextualized problems. Rubrics help students to understand the criteria for judgment from the beginning of their instruction (Montgomery, 2002). As tasks become more complex, there is often a gradual degradation of the structure and comprehension of the rubric. This problem is manifested especially when analytical rubrics are used (rubrics that break the evaluation down to simple components that are scored separately and then combined to produce the global evaluation). Quality criteria are difficult to use, both by teachers and students, if they become too abstract. A typical approach to rectify this issue is to disaggregate the complex criteria into a series of more understandable criteria of lower conceptual difficulty. A problem arises when a compact list of abstract or dense criteria is replaced by a long list of simpler ones, which in many cases can make them impractical and time-consuming. This situation is accentuated when each criterion is weighted to reflect its relative importance. Current Learning Management Systems (LMS) do not provide a solution to this problem, and as a result, analytical rubrics are often avoided in complex situations.

In this context, the concept of an adaptable rubric emerges as a powerful mechanism to support different learning styles and rhythms. We define adaptable rubrics as those that provide multiple levels of detail, which can be expanded on demand. The level of detail can be adjusted and adapted to a specific teaching scenario and/or the students’ level of understanding of quality concepts. If a student finds a particular criterion or its performance levels too difficult to understand, he/she can deploy an additional level of detail (if provided) for that specific criterion, where it is divided into several sub-criteria with a lower abstraction level.
In this paper, we present a new computer-assisted rubric platform specifically designed to support adaptable rubrics. The main features of this platform are:

- provides feedback (showing detailed scores and levels of performance, if requested).
- supports different learning rhythms and styles (different levels of detail are deployed on demand by students, based on their choice).
- collects metadata that could be used to support adaptive behavior in the future.
- automates the management of different weights among scoring criteria during rubric creation.

The platform is generic, as it can be used to manage any type of rubric. The implementation strategy, validation, and lessons learned while developing and testing our platform are presented in the paper. As an example of the application to a highly difficult and complex assessment problem, the developed system was used in a Mechanical Computer Aided Design (MCAD) training scenario at the undergraduate and graduate college level.

The paper is structured as follows: second section describes the state of the art in platforms for scoring rubrics and confirms the lack of support for adaptable rubrics. Third section describes the architecture of the proposed system. The description includes design specifications and the most relevant implementation details. Two experiments aimed at validating the system are described in fourth section. The first experiment was based on spreadsheet forms while the second was based on our new platform. Results confirm that the new tool does not negatively affect the evaluation process. Furthermore, the first experiment demonstrates the weaknesses of spreadsheet forms while the second experiment demonstrates that the new platform provides richer and more meaningful information than other systems. We conclude by highlighting the lessons learned which will be used to guide future developments and improvements of the platform.

Related work

Many authors claim that rubrics can both teach and evaluate (Andrade & Du, 2005; Jonsson & Svingby, 2007; Reddy & Andrade, 2010). In fact, the formative use of scoring rubrics has proven useful under many circumstances (Panadero & Jonsson, 2013). Unfortunately, the use of rubrics in complex contexts is a non-trivial challenge. For instance, MCAD systems are complex software tools that require not only a thorough understanding of the various functionalities provided by the system, but also the application of efficient strategies to create high quality CAD models. Based on related research, this issue can be considered a particular type of the “content-based constructed responses” problem (Liu et al., 2014).

Computer-assisted grading rubrics are essential to improve the efficiency and effectiveness of grading (Anglin et al., 2008; Auvinen et al., 2009). Although some commercially available Computer Assisted Assessment (CAA) tools provide automatically contextualized feedback (Santos et al., 2009; Cebrían-Robles et al., 2014), and some Learning Management Systems (LMS) provide rubric functionality (Atkinson & Lim, 2013; Isbell & Goomas, 2014), currently available implementations only support static rubrics, which are not flexible in adapting to different learning scenarios.

Building on the general concepts of using rubrics to capture judgment (Mertler, 2001; Karkehabadi, 2013) and the evaluative use of rubrics in higher education (Reddy & Andrade, 2010), this paper focuses on the role of rubrics as instruments for acquiring or reinforcing complex skills (Manson & Olsen 2010; Smit & Birri 2014), assessing engineering questions (McCormick et al., 2015), and supporting the process of formative assessment (Popham, 1997) by advising students about their progress and assisting them in their development (Panadero & Jonsson, 2013). Contrary to holistic rubrics—which only score the overall process or product as a whole without judging the individual components separately (Nitko, 2001)—analytic rubrics are typically used for formative feedback (Mertler, 2001), as they allow the individual scoring of factors—or dimensions—of the product or performance (Moskal, 2000). However, a common problem when introducing performance assessment is evaluating complex competences in a credible way (i.e., whether or not observations of complex behavior can be performed in a reliable and trustworthy manner (Jonsson & Svingby, 2007]). In complex assessment scenarios, when comparing instructor and student judgments, assessment differences can be attributed to the fact that the concepts under evaluation may still be so foreign that students are unable to recognize them (Orsmond, Merry, & Reiling, 1996). This lack of consensus between the instructor assessment and the student self-assessment—i.e., the lack of inter-rater reliability—is useful in detecting problems in the understanding of quality criteria. This type of data can be processed by applying proper statistical analyses to the assessed rubrics (Zaiontz, 2015).
Additionally, we are also interested in metadata that describes how users interact with the platform while completing the rubrics. It has been argued that rubrics should be complemented with “anchors,” i.e., written descriptions, examples that illustrate the various levels of attainment, or work samples (Jonsson & Svingby, 2007). To the best of our knowledge, no results have been reported regarding the use of metadata to discover improved methods to increase the effectiveness of anchors.

Although advanced trainees may not require detailed explanations of each evaluation criterion, more detail may be desirable by novice users. Self-directed learning skills (involving self-assessment and task selection skills) are needed to choose an appropriate on-demand learning pathway (Taminiau et al., 2015). Adaptable rubrics should allow students to display more detail and score low-level criteria on demand. If a particular criterion in the rubric is too abstract or difficult to understand, the additional level of detail can provide a clearer description of the expected performance levels (Company et al., 2015). A major challenge with adaptable rubrics involves guaranteeing consistent scoring for each different combination of detailed-level responses. In this arena, spreadsheets have proven impractical, as its implementation requires significant programming proficiency and extracting relevant information from them is time consuming (Company et al., 2016). Furthermore, they can hardly convey performance criteria descriptors, which are critical components of rubric design (Tierney & Simon, 2004).

To distinguish user-driven rubrics from system-driven rubrics, we adopt the terms adaptable and adaptive. While an adaptable rubric can be modified by the user to adapt to different needs, an adaptive rubric is able to adapt or change itself, depending on the usage pattern. The origins of adaptive rubrics are briefly summarized by Georgiadou et al. (2006).

According to Economides and Roupas (2007), the majority of the Computer Adaptive Testing (CAT) systems “do not offer any advanced support and functionalities to the examinee.” In analyzing current available LMS systems such as Moodle, Sakai, Blackboard or Desire to Learn (D2L), it was found that none of them implement functionality to support adaptable rubrics. They only support static rubrics that cannot deploy additional levels of detail to improve students’ understanding.

**Electronic platform for adaptable rubrics**

This section describes the design and implementation of a dedicated rubrics platform, with support for adaptable rubrics, and the ability to output metadata to analyze the evaluation process to guide further improvements.

**Design specifications**

Apart from supporting adaptable rubrics, the main goal of our design was to centralize both rubrics and assessment results in order to improve data exchange and reduce the time required to program and manipulate a rubric. An additional goal was to make the rubric creation process easier by providing specialized and intuitive tools to enter new rubrics into the system and edit existing ones. Finally, metadata management to support learning analytics was also included. Metadata describing the context of the evaluation process is required to better understand current problems and guide future improvements. We define metadata as the dataset collected by tracking how users interact with the platform while completing the rubrics. Spreadsheet-forms (which cannot collect such data) and proprietary “black-box” rubric platforms (which do not grant direct access to the data) are unsuitable.

Two main specifications were defined for our rubrics platform: (1) the system must be adaptable and allow for rubrics with varying levels of detail, and (2) the rubrics should be easy to reconfigure by allowing instructors to adjust them to changing scenarios.

The first specification implies that every participant that uses the rubric to score a particular task should be able to select the desired level of detail. The rationale behind our adaptable rubrics is the “expand–contract strategy,” which adapts the rubrics to trainee progress and assists them in comprehending the different dimensions of the rubric (Company et al., 2015). Although technically there is not a limit to the number of supported levels, our first prototype has been intentionally limited to three levels. The second specification is intended to allow instructors to customize rubrics based on the evolution of student learning. In addition, our long-term goal involves linking the rubrics to appropriate anchors, as defined by (Jonsson & Svingby, 2007).
Three additional design specifications were also considered: (3) the system must provide immediate feedback of the evaluation scores; (4) instructors should be able to easily extract and process the information of the completed forms, and (5) forms should prevent incomplete or inconsistent scoring.

From a user standpoint, the platform manages three types of accounts: students, instructors, and administrators. After a successful login to the system, students are allowed to perform evaluations using rubrics (both self and peer evaluation). Instructors are allowed to manage rubrics, manage student users and groups, assign rubrics to groups, schedule rubric assignments, and manage results. A global administrator can create new instructors.

Prototype implementation

Before implementing a dedicated rubrics platform, our team tested the concept of adaptable rubrics using a spreadsheet-based mock-up, which was used in a complex assessment scenario, showing promising results regarding students’ improved management and understanding of the rubric (Company et al., 2016). A spreadsheet was configured to support rubrics with multilevel criteria, which could be unfolded using buttons. A five-level Likert scale implemented by radio buttons represented performance levels. Active X controls and Visual Basic for Applications (VBA) were used for the implementation. An example of deployment of sub-criteria is presented in Figure 1. In this case, when level 2 sub-criteria for criterion 1 are unfolded, criterion 1 is blocked, and the student can only mark sub-criteria 1.1 and 1.2. As a result, the score for criterion 1 is automatically calculated and displayed as the total score for this criterion.

Implementation

Spreadsheet-based mock-up showed many limitations to create a general tool to support the creation and use of adaptable rubrics, so our team opted to develop Annota eRubrics, a web-based framework to manage adaptable rubrics and users. Three user profiles were defined: student, instructor, and global administrator. Student users can fill out rubrics and visualize completed rubrics during the evaluation period defined by the instructor. Instructors can manage rubrics (create new, import from CSV or XML files, export to files, and assign rubrics to students during a certain period of time), manage students (register new, import from CSV files, and manage groups of students), and access assessment results (view and export to files). Finally, global administrators (who can also be instructors) can register new instructors.
The general structure of the framework is shown in Figure 2. Data is stored in a MySQL database and accessed via an Apache web server by a number of PHP scripts triggered by the framework front-end running over Unity3D Webplayer (Creighton, 2010). Only name and e-mail address are required to create a new user. During registration, a custom link is automatically emailed to the users so they can set up their password and enter basic demographics information to complete the process.

A screen capture of a sample MCAD rubric with 6 main criteria (for the 6 CAD quality dimensions to be assessed) is shown in Figure 3.

An example of the mechanism to unfold lower level sub-criteria is illustrated in Figure 4. By clicking the “+” (unfold) or “−” (fold) signs located on the upper left corner of the criteria (criteria 1 and 1.2 have been unfolded in Figure 4), users can dynamically adapt the rubric to their own rhythms and learning styles. Instructors can pre-configure criteria as “folded” or “unfolded” to determine which criteria will be viewed folded/unfolded by default when students load the rubric for the first time.

Each criterion has an associated scoring weight configured by the instructor (Figure 5). Assessment results are recalculated every time the student fills out a criterion. By default, weights and assessment results are not visible to the student but can be activated at any time. This activation/deactivation is part of the metadata collected by the system.
As a result of our first experiment (described in the next section), contextual information bubbles were added to various options of the rubric. These bubbles display textual explanations for each level of deployment, also called performance criteria descriptors, and are automatically shown when the user hovers over a check box (Figure 6). These bubbles provide a complementary explanation of the Likert level associated to a specific criterion. The number of bubble activations and the bubble activation times are also part of the collected metadata.

From an instructor standpoint, a “student view” is available, which allows instructors to view the rubric as members of the student group. This mode is for testing purposes only, so results are not saved. In addition, instructors can grade students by importing a CSV file. This functionality allows the use of spreadsheets to grade an entire user group with a single mouse click.

Experiments

To validate our platform, we selected a complex assessment problem in the area of MCAD training. We conducted two experiments with various groups of MCAD students. The first experiment validates adaptable rubrics for feedback and evaluation, but states the limitations of spreadsheet-based adaptable rubrics. It also provides a comparison base to determine that the new form, used in the second experiment, does not negatively affect the behavior of students. The second experiment also shows how our new platform for adaptable rubrics can offer additional guidelines to help evaluation, when required. Finally, the second experiment illustrates a future development in which the metadata generated by the new platform can be used as a resource to identify weaknesses, measure effectiveness, and lead to further improvements.

Experiment 1

The goal of our first experiment is to test the feasibility and analyze the benefits of the adaptable rubrics concept. We use a learning/teaching scenario based on the combination of suitable tutorials and materials during instruction with a set of adaptable rubrics for feedback and evaluation. The context is MCAD training.

Our sample was a multi-disciplinary group of senior engineering students at a US university. Students were divided into two groups: experimental (EG) and control (CG). Both groups were given a complete set of lecture notes, which included detailed explanations of the different dimensions of MCAD quality (we refer to it as theoretical background). The lecture notes given to the EG also included detailed explanations and examples on how to apply rubrics to self-evaluate in-class exercises (practical guidelines). This information was intentionally removed from the lecture notes that were given to the CG.

Participants were asked to create 3D solid models of two parts (see Figure 7). As an additional requirement for the second part, participants were explicitly instructed to make their models flexible and reusable by allowing a series of specific design changes to be performed successfully and efficiently. Changes for the part 2 EG included: Increasing the distance between the cylinders with holes (for example, from 2 to 3 inches), modifying the height of the Ø3/4 cylinder (for example, from 7/8 to 1 inch), and rotating the arm (for example, 90 degrees). Changes for the part 2 CG included: Increasing the distance between the center points of the countersunk holes (for example, from 90 mm to 150 mm), increasing the height of the main cylinder (for example, from 65 mm to 100 mm), and rotating the arm (for example, 90 degrees).
All subjects were asked to self-evaluate their work using spreadsheet-based adaptable rubrics. After submitting the first self-evaluation, they were given a sample “solution:” step-by-step instructions of an efficient strategy (defined by our research team) to model the parts, which also included detailed explanations on how to evaluate the quality of the models. Provided with this information, they were asked to re-evaluate their original models a second time.

The rubric used to evaluate the experiment was necessarily multidimensional, as is the nature of the quality of CAD models. Thus, a multilevel criteria tree was selected. The detailed sub-dimensions schema used for the experiment is shown in the tree diagrams of Figures 8 and 9.

The initial sample size of the study was 50 subjects in the EG and 49 subjects in the CG, but only 29 in the EG and 26 in the CG completed all tasks.

As expected, all the inconveniences of spreadsheet-based forms soon emerged. For example, 8.8% of the rubric forms (29/330) were returned unlocked and evaluated simultaneously at various levels. This issue resulted in inconsistent scores between the main levels and their corresponding sub-levels. Also, the process of extracting the information from the forms was time consuming and prone to errors. The only metadata we could obtain from those forms was that only 7 out of 330 rubrics were assessed without showing the score (thus, the usefulness of such feedback was demonstrated). Finally, by analyzing student submissions, we determined that 53% of the rubrics were always used at the third level (175/330), while 8% of the assessments only deployed the first level (26/330). The remaining assessments used Level 1 for some dimensions and Levels 2 or 3 for others, thus demonstrating that the rhythm of assimilation of quality concepts varies among students and validates the need for adaptable rubrics.

We searched for dissimilarities in the understanding of quality criteria by comparing significant differences between inter-rater evaluations. Results for Part 1 EG and Part 1 CG illustrated similar behavior in both user groups. Figures 10 and 11 visually summarize the detailed quantitative results by showing the differences between the instructor, self, and peer evaluations. Each quality criterion is marked with a grayed background if
the student and the instructor agree within a threshold (the threshold Delta was set to 0.25, which is equivalent to a jump between two consecutive levels in a five-level Likert item mapped into the range \([0, 1]\)), and is void for larger differences.

**CAD model quality**

- **The model is valid**
  - The file of the model can be located and opens in neutral state
    - The file of the model has the expected contents (and name) and is in the expected folder (or website)
    - The file of the model can be re-opened after closing the current session (even on a different computer)
    - The file of the model opens in neutral state (i.e. no operations are in progress)
  - The model can be used
    - Model is compatible with the CAD of the receiver
    - Model tree is free from error messages

- **The model is complete**
  - The model replicates the shape of the part
  - The model replicates the size of the part

- **The model is consistent**
  - Profiles are free from duplicated and segmented lines, and are fully constrained
    - Profiles are free from duplicated lines
    - Profiles are free from segmented lines
    - Profiles are fully constrained
  - The model tree is strongly linked to the global reference system and to a suitable set of datums
    - The model is aligned and oriented with respect to global reference system
    - Model uses suitable datums (that define a scaffold that helps build and edit the model)
  - The model tree is free from unnecessary dependencies
    - Functional elements are defined by independent modeling operations
    - The parent/child relations in the model tree are free of unnecessary dependences

*Figure 8.* Tree structure of the assessment criteria used to evaluate the three first main criteria of CAD model quality
Figure 9. Tree structure of the assessment criteria used to evaluate the three last main criteria of CAD model quality
The Wilcoxon (non-parametric) test for related samples was applied, since normality could not be assumed (Kolmogorov-Smirnov test, at the 5% level of significance). Tables 1 and 2 summarize the significant agreements between the instructor and self and peer evaluations for both the CG and EG (part 1).
differences. In light of those results, the practical guidelines provided to the EG do not seem to have any effect on the results. The detailed instructions to model the parts provided for the second auto-evaluation do not seem useful either, since all student evaluations (first and second auto-evaluation, and peer evaluation) showed significant differences with instructor evaluation. Dimension 5 (Clear) is the only one that did not show any differences between evaluations. There was no improvement in the second self-evaluation with respect to the first one, which suggests that short repetitive exposure to rubrics does not increase reliability.

Table 1. Significant differences between evaluations for the CG (part 1)

<table>
<thead>
<tr>
<th></th>
<th>Valid</th>
<th>Complete</th>
<th>Consistent</th>
<th>Concise</th>
<th>Clear</th>
<th>Design Intent</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor and Self1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Instructor and Self2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Instructor and Peer</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>

*Note. *"p = .01; **"p = .05.

Table 2. Significant differences between evaluations for the EG (part 1)

<table>
<thead>
<tr>
<th></th>
<th>Valid</th>
<th>Complete</th>
<th>Consistent</th>
<th>Concise</th>
<th>Clear</th>
<th>Design Intent</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor and Self1</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Instructor and Self2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Instructor and Peer</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>

*Note. *"p = .01; **"p = .05.

Different results were obtained for Part 2 EG and Part 2 CG. Students were explicitly asked to model the part so it would lend itself to specific design changes. This requirement of flexibility matches various sub-levels of Dimension 6 (Design Intent) in the rubric. As a result of these specific conditions, differences arise in the evaluation of Dimension 6 between the CG and the EG (Table 3). Figures 12 and 13 visually summarize the detailed quantitative results.

Figure 12. Visual map of agreements between the instructor, self, and peer evaluations for the CG (part 2)
While significant differences between instructor and student remained consistent throughout ongoing evaluations for CG, they decreased and eventually disappeared from the EG. We hypothesize that the experience gained in successive evaluations appears to improve student assessment only if the practical guidelines are combined with explicit metrics: the background of the EG (the explanations on how to self-evaluate exercises with rubrics) is boosted by specific checking requirements added to the exercise so that students can clearly identify with the criteria.

Table 3. Significant differences between evaluations of the CG and the EG for dimension 6 (part 2)

<table>
<thead>
<tr>
<th>Instructor and Self1</th>
<th>CG</th>
<th>EG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dim 6 (Design Intent)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Dim 6 (Design Intent)</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>Instructor and Peer</td>
<td>**</td>
<td></td>
</tr>
</tbody>
</table>

Note. **p = .01; *p = .05.

The spreadsheet-based prototype revealed the benefits of adaptable rubrics that provide varying levels of detail for each criterion and immediate feedback of the evaluation scores. Every subject was able to select the level of detail he/she required, while being blocked from incoherently marking rubrics when working at different levels. Unfortunately, an extensive use of the tool confirmed that spreadsheet forms are difficult to reconfigure and adapt; information from completed forms is difficult to extract and process, and spreadsheet forms do not fully prevent incomplete or inconsistent scoring. A more generic and flexible platform was required to maximize the benefits associated to the adaptable nature of the rubric.

Experiment 2

The second experiment confirms that the new platform does not affect the reliability of the evaluations, while providing richer and more complete information, as it outputs user interaction metadata that can be used to analyze the evaluation process and guide further improvements.
Our sample included one group of junior industrial engineering students from a Spanish university. All 47 students had basic knowledge of rubrics (MCAD quality concepts were explained using rubrics, but students were never required to practice).

First, students were asked to create a solid model of a specific part (a fixed arm of a pistol clamp) depicted in a detailed drawing (see Figure 14). Next, they were asked to create their own detailed drawing of the fixed arm they previously modeled. Finally, participants were required to assemble a virtual model of the pistol clamp (see Figure 15). All parts were provided, with the exception of the standard parts (available in the library) and the fixed arm (previously modeled by each subject).

Figure 14. Input (left) and output (right) of the part used in experiment 2

Figure 15. Input (left) and output (right) of the pistol clamp used in experiment 2

Figure 16 visually summarizes the detailed quantitative results of the experiments, as they show the differences between teacher and self-evaluations.
Comparison between visual maps of the Experiments 1 and 2 clearly show that students behave quite similarly with independence on whether they evaluate with spreadsheets or with the new platform; the percentages of agreements are similar for both experiments. Thus, we validate the hypothesis that the new platform does not negatively affect the results.

In order to find significant differences between instructor and students evaluations, a T-test for related samples was applied for the average ratings (since the Kolmogorov-Smirnov test revealed normality could be assumed at the 5% level of significance) and the Wilcoxon (non-parametric) test for related samples was applied for Dimensions 1 to 6 (normal distribution could not be assumed). Results are shown in Table 4.

Some collateral lessons learned resulted from Experiment 2. Information bubbles were incorporated to Annota to contextually display textual explanations for each level of deployment (see Figure 6). In addition, the new platform collects information about how users interact with the platform while completing the rubrics.

![Figure 16. Visual map of agreements between instructor and self-evaluations for the three tasks (modeling, drawing and assembling)](image-url)
particular, the platform tracks how long each detailed explanation was visible, so we can identify students who looked at the bubbles longer than average, and those who looked at them a shorter time.

Table 4. Significant differences levels between self-evaluations and instructor evaluation for experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Valid</th>
<th>Complete</th>
<th>Consistent</th>
<th>Concise</th>
<th>Clear</th>
<th>Design Intent</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling task</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing task</td>
<td></td>
<td>**</td>
<td></td>
<td>**</td>
<td>**</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Assembling task</td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

Note. ""p = .01; """"p = .05.

For all tasks (modeling, drawing, and assembly), students who looked at the bubbles longer than average showed less significant differences between their evaluations and the instructor’s, as shown in Tables 5 and 6. Therefore, using bubbles to provide explanations seems to be a valid anchor to illustrate the various levels of attainment.

Table 5. Significant differences between self-evaluations and instructor evaluation for experiment 2, only for students who looked at the bubbles longer than average

<table>
<thead>
<tr>
<th></th>
<th>Valid</th>
<th>Complete</th>
<th>Consistent</th>
<th>Concise</th>
<th>Clear</th>
<th>Design Intent</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling task</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing task</td>
<td></td>
<td>**</td>
<td></td>
<td>**</td>
<td>**</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Assembling task</td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

Note. ""p = .01; """"p = .05.

Table 6. Significant differences between self-evaluations and instructor evaluation for experiment 2, only for students who looked at the bubbles shorter than average

<table>
<thead>
<tr>
<th></th>
<th>Valid</th>
<th>Complete</th>
<th>Consistent</th>
<th>Concise</th>
<th>Clear</th>
<th>Design Intent</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling task</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing task</td>
<td></td>
<td>**</td>
<td></td>
<td>**</td>
<td>**</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Assembling task</td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

Note. ""p = .01; """"p = .05.

The ability of the platform to collect metrics based on user interaction helps to improve the teaching strategy. For example, by examining metadata about the use of bubbles for the modeling rubric (Table 7), we can see that criterion 6 displayed the highest number of bubbles, which can be related to the fact that criterion 6 is more difficult to understand.

Significant differences between student and instructor evaluations were found in all cases (Tables 4 to 6). Maximum and average times significantly increased for both drawing and assembly tasks (Table 7), which can be attributed to the fact that students were less exposed to rubrics of drawings and assemblies than they were to rubrics of part models. Under these circumstances, and based on Orsmond et al. (1996), we presume that the concepts of drawing and assembly quality are still so alien that many students need more time to read the explanations from the bubbles. More illustrative anchors may still be required.

Table 7. Time (seconds) each detailed explanation was displayed in experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Valid</th>
<th>Complete</th>
<th>Consistent</th>
<th>Concise</th>
<th>Clear</th>
<th>Design Intent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max time (seconds)</td>
<td>38.1</td>
<td>6.958</td>
<td>42.79</td>
<td>9.891</td>
<td>15.57</td>
<td>49.86</td>
<td>139.504</td>
</tr>
<tr>
<td>Average time (seconds)</td>
<td>4.573</td>
<td>1.609</td>
<td>3.497</td>
<td>2.114</td>
<td>2.292</td>
<td>2.906</td>
<td>14.032</td>
</tr>
<tr>
<td>Number of users of bubbles</td>
<td>35</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>38</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>Drawing task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max time (seconds)</td>
<td>14.320</td>
<td>27.672</td>
<td>148.351</td>
<td>44.921</td>
<td>13.195</td>
<td>14.910</td>
<td>161.092</td>
</tr>
<tr>
<td>Average time (seconds)</td>
<td>3.173</td>
<td>2.1</td>
<td>5.689</td>
<td>2.96</td>
<td>2.44</td>
<td>1.836</td>
<td>15.036</td>
</tr>
<tr>
<td>Number of users of bubbles</td>
<td>30</td>
<td>43</td>
<td>42</td>
<td>39</td>
<td>36</td>
<td>43</td>
<td>47</td>
</tr>
<tr>
<td>Assembling task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max time (seconds)</td>
<td>33.565</td>
<td>139.501</td>
<td>46.807</td>
<td>61.097</td>
<td>19.988</td>
<td>30.920</td>
<td>285.859</td>
</tr>
<tr>
<td>Average time (seconds)</td>
<td>3.967</td>
<td>6.355</td>
<td>3.006</td>
<td>4.865</td>
<td>2.937</td>
<td>3.003</td>
<td>22.549</td>
</tr>
<tr>
<td>Number of users of bubbles</td>
<td>40</td>
<td>46</td>
<td>44</td>
<td>43</td>
<td>44</td>
<td>46</td>
<td>47</td>
</tr>
</tbody>
</table>

Our Annota platform overcomes a number of disadvantages commonly found in spreadsheet forms. For example, it prevents duplicate evaluations of the same criterion at multiple levels, and provides for easily obtainable metadata. Additional improvements have been gradually incorporated, such as detailed explanations for each
level of deployment that can be shown during the evaluation process, if required. Our e-rubric platform can also be helpful to test the effectiveness of modeling guidelines and other materials, particularly if we are able to use the metadata provided by the platform to detect the materials that better reduce the inter-rater agreement.

Based on our results, we can confirm that the metadata captured by our rubrics platform provides valuable metrics (such as the interaction between the user and the anchors). The hypotheses that was postulated as a result of the metadata analysis will be explored in future experiments. If the hypotheses are true, further work will be required to more accurately correlate such complex information and obtain guidelines to improve the anchors used to recognize quality criteria.

Conclusions and future developments

In this paper, we have tested our concept of adaptable rubrics, and have described the design, implementation, and testing of a web-based adaptable rubrics system aimed at formative e-assessment. It has been proven valid, as demonstrated in a case study on MCAD training that encompasses multiple and heterogeneous criteria to assess, and exhibited a moderately good correlation between the self-assessment of the students, the peer-assessment, and the assessment provided by the instructors. The main novelty of the approach is its adaptability, as it allows users to dynamically access multiple levels of detail for each quality criterion. In our implementation, buttons that allow users to show or hide details, as needed, control the level of detail. In addition, the system outputs metadata, which is helpful to analyze the evaluation process.

Comparing Experiment 1 (where students used spreadsheet forms) and Experiment 2 (where the new adaptable rubrics system was used), we conclude that the new form does not negatively affect the behavior of students, as inter-rater evaluations are quite similar for both experiments, and the small differences seem to have other plausible causes than the difference in the form. Besides, we can speculate that our new platform supports adaptable rubrics that work on some multidimensional rubrics with non-trivial evaluation criteria, where the available literature states that other rubrics do not work accurately.

Some collateral lessons gleaned from the experiments include: (1) specific instruction on the quality concepts evaluated by the rubrics appears to be useful only if the evaluated concepts are directly linked to quantitative metrics; (2) complementary information automatically displayed during the interaction with the rubrics platform seems to have a beneficial effect, and (3) the capability to collect metadata looks promising, but still requires further development to improve the educational process.

Our platform lays the foundation for a future intelligent tutoring system based on adaptable rubrics, which is hoped to easily become also adaptive by establishing a set of interconnected tasks—linked to different rubrics—to automatically activate subsequent tasks that can better assist the student in future learning stages. This process is performed after the previous stage has been self-evaluated and the platform has analyzed the level of performance attained by the student.

In the short-term, we are designing a more comprehensive analysis of the metadata captured by the rubrics platform. This analysis will provide guidelines to improve the anchors and help students better understand rubric’s criteria. In the long-term, we plan to use our system to merge our current MCAD teaching notes and materials (which have already been successfully tested separately) with self-evaluation rubrics in order to develop an online MCAD course.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund, through the ANNOTA project (Ref. TIN2013-46036-C3-1-R). The authors also wish to thank the editor and reviewers for their valuable comments and suggestions that helped us improve the quality of the paper.

References


Differences in Views of School Principals and Teachers regarding Technology Integration

Magdalena Claro¹-²*, Miguel Nussbaum³-⁴, Ximena López⁴ and Victoria Contardo³

¹Faculty of Education, Universidad Católica de Chile, Vichuquén Mackenna 4860, Macul, Santiago, Chile  /  ²CEPPE-UC, Universidad Católica de Chile, Chile  /  ³Escuela de Ingeniería, Universidad Católica de Chile, Chile  /  ⁴School of Computing and Engineering, University of Huddersfield, Queensgate, Huddersfield, United Kingdom  /  mclarot@uc.cl / mn@ing.puc.cl / ximelopez@gmail.com / victoria.contardo@gmail.com

*Corresponding author

(Submitted December 30, 2015; Revised July 1, 2016; Accepted August 9, 2016)

ABSTRACT

This paper studies the similarities and differences among the views of school principals and teachers regarding a mobile computer lab (MCL) initiative implemented in 1,591 public schools in Chile. It also characterizes the aspects in which their views diverge. A mixed methods study was carried out in two stages: first, a quantitative stage, where a self-administered (web-based) questionnaire was sent to the schools; and second, a qualitative stage, where a case study was conducted with three schools. The results show a greater convergence of the teachers’ and school principals’ views regarding the contribution of ICT resources to teaching, with more divergence when it comes to the implementation process. More specifically, these differences were related to two points: (1) how appropriate the conditions were for using and learning how to use the new resources within the context of the school, and (2) who should be held accountable for integrating ICT resources within the school organization. Furthermore, the qualitative results revealed that school principals only had vague information on the pedagogical integration of the MCLs in their schools. These findings suggest that in order to have more effective technology integration processes in schools, a closer presence of school leaders in the teachers’ everyday pedagogical activities is required.

Keywords

Technology integration, Mobile computer lab, Teacher, Principal

Introduction

Any strategy that seeks to change the teaching practice should consider the social and cultural context of the school organization (Hargreaves, Earl, Moore, & Manning, 2001; Tondeur, Devos, Van Houtte, Van Braak, & Valcke, 2009). This means taking into account sociocultural aspects relating to the knowledge, meanings and understanding of the new strategy by the members of a school organization, as well as the changes in social relations it may produce (Cooper, 1988). One common issue when implementing new strategies with ICT is that they tend to focus on adopting the technology, without providing the appropriate conditions for the social and cultural learning that is required for innovation (Hargreaves et al., 2001). Among these conditions, a shared view by the school members that are involved is essential. This shared view includes their perceptions and beliefs of the new strategy that is to be adopted, as well as the physical, human, and organizational conditions required for implementation (Alghamdi & Prestridge, 2015).

School principals’ and teachers’ perspectives

Every organizational change deals with different sub-cultures or a diversity of interests and perspectives that influence the processes and practices of schooling (Ertmer & Ottenbreit-Leftwich, 2010; Prestridge, 2012). When integrating a new technology, the school members’ beliefs and attitudes regarding the use of technology have a direct impact on the integration (Alghamdi, & Prestridge, 2015; Howard, Chan, & Caputi, 2015). As Selwyn (2011) states, there are two relevant sub-cultures in schools: the administrative and the academic. Each one of these has a different logic and way of influencing and perceiving the school processes (e.g. technology adoption).

For school administrators or principals, the predominant logic is one of efficiency. These stakeholders are vital for creating the necessary conditions for a school reform to be successful (Hargreaves et al., 2001; Neufeld, Dong, & Higgins, 2007). Evidence shows that school principals who support and lead teachers when integrating technology into their practices have a clear vision of how the technology will contribute to the school project (Chang, 2012). Their involvement is vital for the technology to be sustainable over time, notwithstanding the amount of teacher training (Law, Pelgrum & Plomp, 2008; Peled, Kali, & Dori, 2011).
On the other hand, teachers’ beliefs regarding the nature of the teaching and learning process is directly related to technology integration (Kim, Kim, Lee, Spector, & DeMeester, 2013). In this sense, the use of technology is motivated by the belief that ICT can help them achieve their pedagogical objectives (Ertmer & Ottenbreit-Leftwich, 2010). Consequently, any new ICT strategy within a school should consider addressing teachers’ beliefs and ideas (Buabeng-Andoh, 2012; Chen, 2010; Ertmer, Ottenbreit-Leftwich, Sadik, Sendurur, & Sendurur, 2012; Mama & Hennesy, 2013).

The importance of a shared vision among school members

Having a shared vision and an ICT policy plan are essential conditions for technology integration (Alghamdi, & Prestridge, 2015). A difference in the views of the school principal and the teachers can be a significant obstacle when it comes to implementing public policy and strategies. In fact, perceptions of how useful an innovation is to professional practice are as essential to its success as the usefulness of the innovation itself (Kirkland & Sutch, 2009; Prestridge, 2012; Shin, 2015; Teo, Milutinović, & Zhou, 2015). These perceptions may be built on the teacher’s beliefs and attitudes, as well as on the influence of the school principal (Kirkland & Sutch, 2009).

The vision and understanding that the school principal may have of the role of ICT in their school should translate into concrete measures to provide teachers with the space to learn how to effectively use ICT in the classroom (Kim et al., 2013; Law et al., 2008). For example, providing the necessary time and support for teachers to prepare to use the technology, research digital materials for their classes, and become familiar with the hardware and software (Jones, 2004). In general, an atmosphere that supports innovation and the use of ICT encourages teachers to try out new practices (Kirkland & Sutch, 2009). In order to do so, it is necessary to implement methods of professional planning and learning that are well-integrated with the teaching process (i.e. not just an accessory), and where learning to teach becomes part of the teaching process itself (Hargreaves et al., 2001; Kim et al., 2013).

In summary, research indicates that the beliefs and understandings of school principals and teachers shape classroom practices. Furthermore, it also reveals the importance of having a shared vision of the usefulness of a new technological strategy, as well as the conditions for integrating the strategy into school practices. Consequently, identifying the teachers’ and principals’ perceptions of a new technological proposal can lead to more effective strategies for integrating new technologies in schools. This paper presents a study elaborating on the views of principals and teachers with regards to a new technological proposal. In order to do so, we consider the implementation of a classroom initiative in Chile that had very limited adoption (Claro, Nussbaum, López, & Díaz, 2013). The research questions are:

- How similar or different are the views of school principals and teachers regarding a mobile computer lab strategy?
- Where do the teachers’ and school principals’ views of a mobile computer lab diverge?

Methodology

The aim of the research was to study and compare the views of principals and teachers regarding the implementation of a new strategy developed in Chile based on a Mobile Computer Laboratory (hereafter, MCL).

The mobile computer lab strategy

This intervention consisted of promoting new classroom practices by providing state primary schools with a cabinet containing laptops loaded with productivity tools. The total number of devices depended on the number of students per class in each school, so that each student in third grade had access to a device. The strategy also included one laptop per teacher, with software to control the class and communicate with students, as well as wireless network technology (intranet) to allow computers within the classroom to communicate (Ministry of Education, 2009). In addition to the technology, a website was also set up to provide information on the project, as well as digital resources in Mathematics and Language Arts to support the teachers’ lessons.

A mixed methods study was carried out in two stages.
Stage one: Survey

In the quantitative stage, an online, self-administered questionnaire was used to study the teachers’ and principals’ views regarding the adoption of the MCL project within their school. The study’s sampling frame consisted of 1,591 schools that participated in the MCL project. Stratified random sampling was applied, using the criteria of Region and rurality to form the required strata. A probabilistic sampling of schools was then applied to each stratum in order to maintain the proportion of schools in each stratum and guarantee the representativeness of the sample. A total of 565 schools were contacted through three successive calls. An e-mail with the URL to access the survey was sent to each school and only one representative from the school could answer the survey. The questionnaire was answered by a total of 242 schools (teachers, school principals, Head of Curriculum and Instruction, or Head of ICT), with a 75% response rate. Given the schools that did not respond, weightings were used to reconstruct the representativeness of the various segments based on their relevant weighting within the overall distribution (for more details, please see Claro et al., 2013). To study the differences or similarities between the views of the school principals and teachers, a chi-squared analysis was conducted using the responses from the school members regarding the main dimensions of the two strands of this study:

- Evaluation of the contribution and adoption of ICT within the school. Specifically, their views on: (a) the contribution of ICT in general; (b) the contribution of the MCL in particular; and (c) the level of pedagogical adoption of ICT within the school.

- Implementation and use of the MCL in the classroom. Specifically, their views on: (a) planning and coordinating the use of the MCL; (b) level of teacher preparation for making pedagogical use of the MCL; (c) technical and pedagogical training for teachers in the pedagogical use of the MCL; and (d) innovative practices and use of the MCL in the classroom.

The specific conditions that were surveyed were selected based on previous research relating to key human and organizational conditions for integrating new technologies in the classroom. Specifically, time for teacher preparation (Jones, 2004), technical and pedagogical support (Kirkland & Sutch, 2009; Law et al., 2008), the school director’s support and leadership (Law et al., 2008) and the school’s ICT plans and strategies (Tondeur, van Keer, van Braak, & Valcke, 2008; Vanderlinde, Aesaert & van Braak, 2014).

Stage two: Case studies

In the second stage, a multiple case study was carried out (Creswell, 2007). Two interviews were conducted at three schools, one with the principal and the other with the teacher. The aim was to further elaborate on the views of each of them and identify areas of convergence and divergence, as well as the difficulties they faced when implementing the MCL strategy in the classroom. For this purpose, and based on the data collected through the questionnaire, three schools were randomly selected: one from the group of schools that reported high use of the MCL, one from the group of schools that reported medium use of the MCL, and one that reported low use of the MCL. This selection criterion was based on the hypothesis that the frequency of use reported by the school may be related to the convergence of the views between school principals and teachers regarding the school’s experience with the MCL. Subsequently, the school principal and a teacher taking part in the project were interviewed at each school. The interviews were conducted following a single set of open-ended questions regarding their perceptions of the same core strands and dimensions included in the quantitative stage: (a) evaluation and adoption of ICT within the school, and (b) implementation and use of the MCL.

The interviews were conducted individually at each school. The principals were interviewed in their offices, whereas the teachers were interviewed in a classroom. Three researchers analyzed the interviews, so as to compare the information and triangulate the data. The interviews were analyzed following a process based on grounded theory (Corbin & Strauss, 2008). In the first stage, the information from the interviews was broken down into units of meaning, which were coded and classified for each of the topics covered. The information was coded using categories for each topic, which in turn were summarized in order to reach a single definition expressing the content of various semantic contributions at three different levels within the discourse that emerged from the interviewees:

- A first level, relating to each interviewee’s general perception of the topic.
- A second level, relating to the in-school or out-of-school organization and/or material conditions that may favor or hinder the development of the topic.
- A third level, relating to those responsible inside or outside of the school for implementing a particular strand.
In a second stage, the categories stemming from each topic at the three levels of discourse for each school member were compared by school. The aim of this was to identify the similarities and differences between the views of the principal and the teacher.

The following cases were studied:

- **Low Use School**: This school serves 381 pre-school, primary school, and secondary school students, of which 80% come from families classified by the Ministry of Education as vulnerable. According to the Ministry of Education’s classification, the school’s level of technology was Elementary, which implies basic infrastructure and precarious pedagogical use of technology. The person in charge of implementing the project, appointed by the school principal, was the Head of Curriculum and Instruction, who organizes and plans all MCL activities. One peculiarity of this school is that the person responsible for teaching the lessons is the Head of ICT, who has to agree on the content and objectives of the MCL classes with the corresponding classroom teacher.

- **Medium Use School**: This school serves 208 kindergarten and primary school students, of which 75% come from families classified as vulnerable. Their level of technology is Advanced according to the Ministry of Education, which implies they have excellent technological infrastructure and pedagogical use is very frequent. At the time of the interviews, the school principal had been in the position for two years and arrived following the implementation of the MCL. A pro-ICT culture is promoted throughout the school, as witnessed in the posters on the walls of the computer room, encouraging the use of computers. The computer room is managed by a teacher who coordinates the timetables and proposes class content to teachers using the MCL. The school premises are seen to be very clean and tidy, which contrasts with the school’s surroundings.

- **High Use School**: This school serves 639 kindergarten and primary school students, of which 60% come from families classified as vulnerable. According to the Ministry of Education, the level of technology is also Advanced. At the time of the visit, the school Principal and the Head of Curriculum and Instruction had been in their positions for two years. The principal mentioned that many ICT projects had not been implemented on account of problems with the school’s internal organization. The school’s Head of ICT provides the MCL on request, but does not give pedagogical support to teachers. In general, there is a positive attitude towards ICT within the school.

## Results

### Stage one: Survey

**Evaluation of the contribution and adoption of ICT within the school**

The chi-squared analysis shows there are no statistically significant differences between the views of the school principals and teachers in any of the types of contributions consulted (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Teachers ((N = 38))</th>
<th>Principals ((N = 53))</th>
<th>(\chi^2)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT are relevant to teaching</td>
<td>94.1%</td>
<td>94.3%</td>
<td>(\chi^2 = 2.93)</td>
<td>0.394</td>
</tr>
<tr>
<td>MCLs are an opportunity to improve digital literacy</td>
<td>74.5%</td>
<td>74.3%</td>
<td>(\chi^2 = 0.001)</td>
<td>0.978</td>
</tr>
<tr>
<td>MCLs are an opportunity to promote pedagogical innovation</td>
<td>92.2%</td>
<td>88.6%</td>
<td>(\chi^2 = 0.425)</td>
<td>0.515</td>
</tr>
<tr>
<td>MCLs are an opportunity to motivate students</td>
<td>88.2%</td>
<td>75.7%</td>
<td>(\chi^2 = 3.005)</td>
<td>0.083</td>
</tr>
<tr>
<td>Would recommend adopting MCLs</td>
<td>94.1%</td>
<td>97.1%</td>
<td>(\chi^2 = 0.682)</td>
<td>0.409</td>
</tr>
</tbody>
</table>

With regards to the adoption of ICT in general, the chi-squared analysis shows there are no statistically significant differences between the views of school principals and teachers (Table 2). In both groups, the majority of respondents reported that they either agreed or strongly agreed that these resources are starting to be used to teach school subjects.
Table 2. Percentage of teachers and principals who say they agree or strongly agree that ICT are starting to be used to teach school subjects

<table>
<thead>
<tr>
<th></th>
<th>Teachers (N = 38)</th>
<th>Principals (N = 53)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agree or strongly agree that ICT are starting to be used to teach school subjects</td>
<td>95.1%</td>
<td>94.4%</td>
<td>$R^2 = 1.52$</td>
</tr>
</tbody>
</table>

Implementation and use of MCLs

With regards to planning and training for the use of MCLs, the chi-squared analysis reveals statistical differences between the views of teachers and principals (Table 3). A significantly higher percentage of principals report that teachers are provided with support in preparing their classes and time for training in how to use the MCL in their subject.

Table 3. Percentage of teachers and principals who report that the following administrative measures are in place in their school

<table>
<thead>
<tr>
<th></th>
<th>Teachers (N = 38)</th>
<th>Principals (N = 53)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigning time for teachers to plan ICT-based classes</td>
<td>45.7%</td>
<td>74.3%</td>
<td>$\chi^2 = 9.76$, $p = 0.002$</td>
</tr>
<tr>
<td>Scheduling time for teacher training to improve the use of the MCL in their subject</td>
<td>65.3%</td>
<td>85.3%</td>
<td>$\chi^2 = 6.40$, $p = 0.011$</td>
</tr>
</tbody>
</table>

There are significant differences when it comes to the views of teachers and school principals regarding technical and pedagogical support (Table 4). In this sense, teachers have a significantly more critical view than principals regarding the frequency with which such support is provided.

Table 4. Percentage of teachers and principals who say the following practices always or never take place at the school

<table>
<thead>
<tr>
<th></th>
<th>Teachers (N = 38)</th>
<th>Principals (N = 53)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never receive pedagogical support from school authorities</td>
<td>40%</td>
<td>10%</td>
<td>$\chi^2 = 17.39$, $p = 0.001$</td>
</tr>
<tr>
<td>Principal never visits MCL classes</td>
<td>37.5%</td>
<td>11.6%</td>
<td>$\chi^2 = 12.54$, $p = 0.06$</td>
</tr>
<tr>
<td>Always receive pedagogical support from school authorities</td>
<td>14%</td>
<td>37%</td>
<td>$\chi^2 = 10.63$, $p = 0.014$</td>
</tr>
<tr>
<td>Always receive technical support from ICT Coordinator</td>
<td>26.5%</td>
<td>54.3%</td>
<td>$\chi^2 = 10.31$, $p = 0.016$</td>
</tr>
</tbody>
</table>

With regards to using MCLs, there are significant differences between the views of teachers and school principals (Table 5). Principals have a statistically significant more positive view than teachers on this topic. Nevertheless, when it comes to adopting new practices with ICT and MCLs, there were no statistically significant differences between the two groups.

Table 5. Percentage of teachers and principals who say that MCLs are used in the classroom to innovate

<table>
<thead>
<tr>
<th></th>
<th>Teachers (N = 38)</th>
<th>Principals (N = 53)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers in the school use MCLs in the classroom</td>
<td>71%</td>
<td>88.4%</td>
<td>$\chi^2 = 8.120$, $p = 0.044$</td>
</tr>
<tr>
<td>Internet is used in the school to support the teaching of different subjects</td>
<td>81.6%</td>
<td>94.1%</td>
<td>$\chi^2 = 4.49$, $p = 0.034$</td>
</tr>
<tr>
<td>Agree or strongly agree that the school is involved in adopting new pedagogical practices using MCLs</td>
<td>94.1%</td>
<td>98.5%</td>
<td>$\chi^2 = 1.62$, $p = 0.219$</td>
</tr>
</tbody>
</table>

Stage two: Case studies

The results from the previous section revealed a convergence of opinions among principals and teachers with regards to their perception of the contribution of ICT and MCLs to teaching (Table 1), their adoption in teaching (Table 2), and the development of new pedagogical practices using MCLs within the school (Table 5). Nevertheless, when analyzing the questions related to the process of implementation, statistically significant
differences appear. More specifically, this relates to the conditions for planning the use of MCLs (Table 3), teacher support and preparation for using MCLs (Table 4) and the teachers’ use of technology in class to support the teaching of school subjects (Table 5).

To understand these results in more detail, a case study was developed to learn about the views and beliefs of teachers and principals regarding the topics under analysis. As mentioned in the Methodology section, the final categories for each topic are presented at the three different levels of discourse that emerged from the analysis of the interviews:

- A first level, relating to each school member’s general perception of the topic.
- A second level, relating to the in-school or out-of-school conditions that may favor or hinder the development of the topic.
- A third level, relating to those responsible inside or outside of the school for implementing the relevant strand.

**Evaluation of the contribution and adoption of ICT within the school**

In terms of the general perception of the contribution of technology, the teachers and principals had a similar perception that ICT in general, and MCLs in particular, contribute to education by motivating the students. With regards to the specific contribution of MCLs to learning, there was also a general agreement within each school (Table 6). This topic was only addressed at the general perception level since it aimed to elicit the interviewees’ general ideas and opinions on ICT and MCLs, which did not directly involve the school conditions and responsibilities, unlike the other topics.

| Table 6. Views of principals and teachers regarding the contribution of ICT and MCLs to teaching |
|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
| General perception                                           | Low use school                                               | Medium use school                                             | High use school                                               |
|                                                              | Principal Teacher                                            | Principal Teacher                                            | Principal Teacher                                            |
| Contribution of ICT                                          | Student motivation                                           | Better outcomes in learning and motivation                   | Student motivation                                           |
| Contribution of MCLs                                         | Access to digital culture and review of school subjects      | Subject learning and digital skills                          | Significant learning and digital skills                       |
|                                                              | Review of school subjects                                    |                                                               | Significant learning and student motivation                  |

**Table 7. Views of principals and teachers regarding the adoption of ICT and MCLs in teaching**

| Table 7. Views of principals and teachers regarding the adoption of ICT and MCLs in teaching |
|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|
| Level of discourse                                           | Low use school                                               | Medium use school                                             | High use school                                               |
|                                                              | Principal Teacher                                            | Principal Teacher                                            | Principal Teacher                                            |
| General perception                                           | Some teachers still do not adopt them                         | Smooth adoption, no resistance to change                     | Adoption is ongoing, some teachers more than others           |
| Organization and/or material conditions (obstacles)           | Need for change in teachers’ culture                         | Need for change in school culture and processes              | Need for technical support for new teaching practices         |
| Responsibilities inside and outside the school                | School leaders and teachers                                   | School leaders and teachers                                   | Teachers                                                      |
|                                                              |                                                               |                                                               | Teachers and Head of Curriculum and Instruction              |
Principals and teachers had similar views in terms of their general perception, the organization and/or material conditions of ICT adoption in teaching, and the school responsibilities (Table 7), with some differences when it comes to their perception of the conditions. In this case, the principals presented a broader view of the organization (e.g. teacher culture, school processes), whereas the teachers approached the topic in more material terms and from the classroom perspective (e.g. need for technical support to change teaching practices).

**Implementation and use of MCLs**

Although there is some convergence in terms of their general perception, when it comes to describing the organization and/or material conditions (obstacles) and responsibilities in formal planning, the views of the principals and teachers are completely divergent (Table 8).

<table>
<thead>
<tr>
<th>Table 8. Views of principals and teachers regarding lesson planning for MCLs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level of discourse</strong></td>
</tr>
<tr>
<td><strong>Principal</strong></td>
</tr>
<tr>
<td><strong>General perception</strong></td>
</tr>
<tr>
<td><strong>Organization and/or material conditions (obstacles)</strong></td>
</tr>
<tr>
<td><strong>Responsibilities inside and outside the school</strong></td>
</tr>
</tbody>
</table>

When it comes to preparing the teachers for using the MCLs, the views are divergent at every level, with the exception of the general perception within the low use school (Table 9).

<table>
<thead>
<tr>
<th>Table 9. Views of principals and teachers regarding the preparation for teachers to use the MCL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level of discourse</strong></td>
</tr>
<tr>
<td><strong>Principal</strong></td>
</tr>
<tr>
<td><strong>General perception</strong></td>
</tr>
<tr>
<td><strong>Organization and/or material conditions (obstacles)</strong></td>
</tr>
<tr>
<td><strong>Responsibilities inside and outside the school</strong></td>
</tr>
</tbody>
</table>

When it comes to organizing school support in order for teachers to use the MCLs, the views of the principals and teachers diverge at every level, with the exception of the general perception within the low use school (Table 10).

In terms of innovative practices, the general perception of the principals and teachers are divergent in the sense that they provide different definitions of the changes and new practices that are expected from the MCL (Table 11). For the conditions that are provided to support innovative practices, the principals and teachers from the low use and medium use schools have divergent views, which is not the case in the high use school. Finally, when it comes to school responsibilities, the teachers and principals from the low and medium use schools have
convergent views, which is again not the case in the high use school (the director and teacher assign each other the responsibility).

| Table 10. Views of principals and teachers regarding the support provided to teachers for using MCLs |
|----------------------------------|------------------|------------------|------------------|------------------|------------------|
| Level of discourse               | Low use school   | Medium use school | High use school  |
|                                  | Principal        | Teacher          | Principal        | Teacher          | Principal        | Teacher          |
| General perception               | There is no formal organization. However, there is support among teachers | There is no formal organization. However, there is support among teachers | There is formal organization | There is informal organization | There is no coaching | Coaching is provided by the ICT Coordinator |
| Organization and/or material conditions (obstacles) | Temporary difficulty (earthquake damage) | Permanent difficulty (no formal support) | Individual formally appointed by the principal (not an obstacle) | Informal help among teachers | Ministry does not provide the necessary conditions | Teacher support is organized by the ICT Coordinator |
| Responsibilities inside and outside the school | External, earthquake | Internal, organization | ICT Coordinator | Principal | Ministry of Education | ICT Coordinator |

| Table 11. Views of principals and teachers regarding innovative practices in the classroom using MCLs |
|----------------------------------|------------------|------------------|------------------|------------------|------------------|
| Level of discourse               | Low use school   | Medium use school | High use school  |
|                                  | Principal        | Teacher          | Principal        | Teacher          | Principal        | Teacher          |
| General perception               | There has not been any innovation, and therefore profound change is needed | There has been some innovation in the sense of new strategies that break up the routine | There has been some innovation in the sense that the technology has been adopted | There has been some innovation in the sense of doing things differently to motivate students | Teachers have not changed their practices | I have changed my practices, but I don’t know about the others |
| Organization and/or material conditions (obstacles) | Need for pedagogical integration of MCLs, teachers need to change | More technology will bring more innovation | Conditions are provided to learn about new practices (not an obstacle) | Time to learn about new practices | Need for sharing new practices | Need for sharing new practices |
| Responsibilities inside and outside the school | Teachers | Teachers | Teachers | Teachers | Teachers | School Principal |

Analysis of results

With regards to the first research question, “How similar or different are the views of school principals and teachers with regards to a mobile computer lab strategy?”, Table 12 provides a summary of the main quantitative and qualitative results in terms of the convergence and divergence of views among school principals and teachers in the two main strands studied. In general terms, the quantitative and qualitative data are consistent in showing a greater convergence of views among teachers and principals for the first strand (i.e. evaluation of the contribution and adoption of ICTs within the school), than the second strand (i.e. the process of...
implementing and using MCLs in the school). More specifically, for the first strand both stages of the study revealed a convergence of views when it comes to the contribution of ICT in general, and MCLs in particular, as a learning resource. The qualitative study also revealed that the evaluation is mostly linked to the resources’ ability to motivate students and assist with the revision of school subjects, as well as developing digital skills. Furthermore, both stages of the study showed a convergence of views among principals and teachers with regards to the adoption of ICT for teaching within their schools. The qualitative study revealed that principals referred to this topic from a broader viewpoint (i.e. based on the organization as a whole), while teachers had a more specific perspective (i.e. based on the classroom and material conditions). However, these views were not contradictory.

In terms of the second strand, regarding the implementation and use of MCLs, both the quantitative and qualitative data revealed divergent views among principals and teachers. The quantitative data revealed statistically significant differences (with a \( p < .05 \)) for topics related with planning time and preparing and supporting teachers in the use of MCLs. The qualitative data, meanwhile, revealed that the views of the principals and teachers were generally divergent with regards to these topics. This is particularly the case with the in-school or out-of-school conditions that may favor or hinder the development of a particular strand and the responsibilities inside or outside the school for its implementation. The only strand where quantitative and qualitative data were contradictory was in the use and innovation of teaching practices in the classroom using MCLs. Although the vast majority of teachers and principals on the survey agreed that new pedagogical practices using ICT were being adopted within their schools, the qualitative data revealed differences in their definition of innovation. In this sense, the principals and teachers did not converge on a concrete view of what they intended to achieve with the new resource in the classroom, nor how they expected to achieve it.

Table 12. Convergence (C) and divergence (D) of the views of principals and teachers – Summary of qualitative and quantitative results

<table>
<thead>
<tr>
<th>Discourse level</th>
<th>Low use</th>
<th>Medium use</th>
<th>High use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contribution of ICT and MCLs</td>
<td>C’</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Adoption of ICT for teaching</td>
<td>C’</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>1. Perception</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>2. Conditions</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>3. Responsibilities</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Teacher preparation for using the MCL</td>
<td>D’</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>1. Perception</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>2. Conditions</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>3. Responsibilities</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>School support for using MCLs</td>
<td>D’</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>1. Perception</td>
<td>C</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>2. Conditions</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>3. Responsibilities</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>Use and innovation with MCLs</td>
<td>C’</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>1. Perception</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>2. Conditions</td>
<td>D</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td>3. Responsibilities</td>
<td>C</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

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

With regards to the second research question, “Where do the teachers’ and school principals’ views of a mobile computer lab diverge?” the main results are summarized below.

First, the analysis reveals that the teachers’ and school principals’ views diverged on two main points: (1) how appropriate the conditions were for using and learning how to use the new resources within the context of the school, and (2) who should be held accountable for integrating ICT resources within the school organization (qualitative data).
With regards to the first point, both the quantitative and qualitative results showed that, in general, teachers had a more negative view than principals on a series of issues. These issues included the conditions for planning the use of the new technology (e.g. formal planning time), the length and quality of the training they received, and the formal technical and pedagogical support they received when using the MCL in the classroom. Although the quantitative data revealed similar views in terms of the use and adoption of new practices using MCLs, the qualitative data showed that the views diverged when it came to defining the expected innovation and the time and conditions required for learning about new practices.

In terms of who should be held accountable within schools for integrating ICT, the qualitative data showed that teachers and school directors, assigned different responsibilities for most of the topics within the implementation and use strand, suggesting that this has not been formally defined. The data also revealed that the school principal did not have a specific leadership role when it came to the use of the MLC within the organization.

As is consistent with the role played by each member of the school, the analysis also revealed that the school principal had a broader approach to most of the topics. This is presumably related to the fact that they are responsible for managing the school organization as a whole. In contrast, the teacher had a more specific approach, based on their responsibility for managing the resources and learning in the classroom. Although this was to be expected, analysis of the interviews also showed that principals were not very involved in the process of implementing and adopting the new technology in the classroom. More specifically, they did not have any concrete information on the type of activities that were carried out, the difficulties encountered, or the results obtained with the students. In general, their perspective on the school conditions was broader and more positive than the teachers’ perspective.

Furthermore, the qualitative data showed that the interviewees’ view of the contribution of MCLs to learning was quite vague. In this sense, the contribution of MCLs tended to be identified with ICT in general (e.g. student motivation, digital skills etc.), failing to see the specific contribution of students having 1:1 access to technology in the classroom.

Finally, no relationship was found between the reported level of use of MCLs (High, Medium, and Low) in the survey and the characteristics of the respondents’ views. This is probably because, in pedagogical terms, the use of the resource in all three cases was neither relevant nor innovative.

**Conclusions**

This research aimed to understand how aligned the views of teachers and school principals were with regards to a new ICT strategy, as well as characterizing the aspects in which their views diverged. In order to do so, a mixed methodology design was implemented. The quantitative and qualitative data was consistent in showing a convergence of views among principals and teachers regarding ICT in general, and MCLs in particular, in terms of being well adopted and contributing to the teaching and learning process within the school organization. The divergence of opinions emerged with regards to two main points: (1) how appropriate the conditions were for learning how to use the new resources within the context of the school, and (2) who should be held accountable for integrating ICT resources within the school organization.

The qualitative study also revealed that, despite the views of the teachers and principals being aligned with regards to the value and general contribution of the new ICT strategy, there were two fundamental problems. First, there was a vague notion among all interviewees regarding the specific contribution of the new resource, which led to an absence of pedagogical intentionality. This is an important obstacle since not properly understanding the usefulness of a new strategy or resource greatly limits the results of its implementation (Kirkland and Sutch, 2009). Secondly, the principals and teachers had different views of the conditions that were in place to adequately incorporate the resource into the teaching and learning process. In this sense, there was no shared diagnosis of the gap between the innovation and the current pedagogical practices. There was also no agreement on the resources that are required in order to achieve this innovation (Kirkland & Sutch, 2009).

Furthermore, the qualitative results revealed the absence of leadership by school principals. This was reflected in their lack of knowledge of what happens on a day-to-day basis in the classroom and the difficulties faced in terms of the time needed for planning and learning in their schools. Additionally, principals did not take direct responsibility for the topics included in the interview. This presents another key problem that has already been identified in previous studies, namely the importance of leadership by the principal for the successful implementation of new strategies. This should translate into providing the appropriate conditions for the
implementation, such as providing time for reflection and learning (Law et al., 2008; Jones, 2004). Finally, the interviews at the three schools also revealed the lack of a suitable environment and organization for learning within the schools. This was clear from the interviewees’ discourse regarding the lack of formal organization for planning classes, technical and pedagogical training, and an exchange of practices and innovations, among others.

In summary, the findings of this study show that having an aligned view of the positive contribution of technology is not enough for successful adoption of ICT in the classroom. More specifically, the qualitative and quantitative data was consistent in showing that the views of school principals and teachers were different, particularly with regards to school conditions and responsibilities. Furthermore, the qualitative results also showed that teachers and principals only had a vague notion of the exact pedagogical contribution of the mobile computer labs, while the school principals did not have any specific information on the process of implementing the new technology in their schools. These findings call for a closer presence of the school principal and other members of the school’s administrative staff in everyday pedagogical activities. Having school leaders be more involved in everyday classroom activities would help close the gap between their views and those of the teachers, regarding how to successfully integrate technology into the classroom. In turn, this should lead to the implementation of more effective technology integration processes in schools.

One limitation of this study is the scope of the technological innovation. Future research should consider the introduction of another type of technological innovation. This would allow us to verify whether the pattern of views found among principals and teachers here is repeated or, alternatively, whether these views are specific to the mobile computer lab strategy in this study and/or the country’s school culture.

Acknowledgements

This research was made possible thanks to support from Conicyt CIE01.

References


Synchronized Pair Configuration in Virtualization-Based Lab for Learning Computer Networks

Chaknarin Kongcharoen¹, Wu-Yuin Hwang²* and Gheorghita Ghinea³

¹Department of Computer Science and Information Engineering, National Central University, Taiwan // ²Graduate Institute of Network Learning Technology, National Central University, Taiwan // ³Department of Computer Science, Brunel University, United Kingdom // csnrk@ku.ac.th // wyhwang@cc.ncu.edu.tw // george.ghinea@brunel.ac.uk

*Corresponding author

(Submitted December 31, 2015; Revised July 21, 2016; Accepted August 10, 2016)

ABSTRACT

More studies are concentrating on using virtualization-based labs to facilitate computer or network learning concepts. Some benefits are lower hardware costs and greater flexibility in reconfiguring computer and network environments. However, few studies have investigated effective mechanisms for using virtualization fully for collaboration. Therefore, this study proposed one effective collaboration mechanism—synchronized pair configuration—in a virtualization-based lab and aimed to enhance interaction and collaboration between paired students to help them accomplish networks or systems configuration through one shared, synchronized terminal, which allows them to input commands simultaneously from two computers to accomplish tasks. Meanwhile, they can communicate through a chat window. In the experiment, university students were divided into an experimental group, using a synchronized pair configuration for lab assignments, and a control group, with students completing lab assignments individually. Results indicated that the experimental group significantly outperformed the control group in learning achievement and had more confidence in their work; furthermore, the instructor’s workload was reduced. Finally, findings of interviews and questionnaires revealed that the experimental group considerably enjoyed the proposed mechanism and system and had high motivation to use them. Therefore, synchronized pair configuration in a virtualization-based lab is suitable for learning computer networks.

Keywords

Synchronized pair configuration, Pair programming, Virtualization, Distributed cognition, Learning computer networks

Introduction

Recently, practice labs in computer network courses have become essential for enhancing students’ learning since they can enable students to apply, practice, and verify network concepts in real contexts. In general, network configuration is usually an individual task and is similar to tasks of famous networking certification examinations such as the Cisco Certified Network Associate and the Linux Professional Institute Certification. Each examinee must individually configure a network on these tests. However, implementation of a physical network lab on campus has limitations in terms of the amount of available hardware since hardware requires rapid upgrades, and the number of instructors seems insufficient to assist all students (Williams, 2010; Xu, Huang, & Tsai, 2014). Meanwhile, some studies have revealed other limitations when students work individually to solve problems in computer science; for instance, solo students had less self-confidence in doing assignments (Williams & Kessler, 2002), made more syntax and typing errors while programming (Lui & Chan, 2006), and required more assistance from instructors (Braught, Wahls, & Eby, 2011), thereby spending more time solving their problems (Williams, Wiebe, Yang, Ferzli, & Miller, 2002). In contrast, previous studies found that pairs outperformed individuals in the number of interactions (Williams, Wiebe, Yang, Ferzli, & Miller, 2002) and learning outcomes in computer programming (Layman, Williams, Slaten, Berenson, & Vouk, 2008). In computer network learning, as experienced teachers in computer networks (with more than 10 years teaching experience), we found insufficient networking hardware and a heavy teaching load when students were assigned to perform lab practices individually. Therefore, in our previous study, stand-alone desktop virtual machines (e.g., VMware and Oracle Virtual box) (Dobrilović & Stojanov, 2006; Chen & Tao, 2012) were enhanced, and one distributed virtual machine (e.g., KVM cloud platform) (Hwang, Kongcharoen, & Ghinea, 2014) was proposed for solving the hardware limitation. However, after deep investigation, we found that some students still did not pay attention in lab practices even though they could observe and learn from their peers or group mates; moreover, they did not ask for much help from instructors. Therefore, we tried to discover an effective collaboration mechanism that would strengthen their collaborations in a lab class and thereby reduce the instructors’ load. This study proposed one synchronized mechanism to help paired students to practice learning computer networks in the virtualization system to reduce class teaching effort and to facilitate sufficient networking lab installation.

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@fits.info.
The virtualization system is a group of virtual machines (VMs) deployed on one server, and the VMs contain operating systems (OSs) for operating as networking equipment, such as hubs, networking switches, and routers. The synchronized mechanism is a group assignment, and it requires a pair of users to work simultaneously. Thus each user works in a pair and pays the same level of attention as his/her partner. The mechanism mentioned above is called synchronized pair configuration (SPC) in this study. Through SPC, paired students can input commands simultaneously from two computers to one terminal. SPC enables a pair of students to work simultaneously and reduces instructor effort. In addition, paired students can share ideas and knowledge during lab assignments without much help from instructors. Due to SPC, this study established a virtualization-based lab (VBLab) that allows conduction of labs in the virtualization system. This VBLab’s design is improved from our previous version (Hwang et al., 2014), with a new feature for a web terminal that allows students simultaneously or separately to input commands to VMs. In addition to experiment procedures, university students were divided into two groups, as pairs and as individuals, and each pair accomplished network lab practices together using SPC while each individual accomplished them alone. Consequently, this study compared pair and individual learning and studied differences in student interaction among lab practices and their effects on learning achievement. Furthermore, students’ perceptions of using our proposed system are also investigated.

**Literature review**

This experiment was undertaken with three well known techniques: pair programming and synchronized programming, a virtualization-based lab, and distributed cognition theory. Details regarding theory and technique are as follows:

**Pair programming and synchronized programming**

Canfora, Cimitile, Garcia, Piattini, and Visaggio (2007) and Williams, Kessler, Cunningham, and Jeffries (2000) originally introduced pair programming in the industry. The concept of pair programming consists of two programmers working collaboratively to develop software. One programmer (the driver) has responsibility to generate code, while the other programmer (the navigator) monitors the driver’s work to safeguard against defects and offers suggestions to the driver regarding various components of the overall software process such as design, coding, and testing. The driver and navigator can swap roles, thus balancing the workload. Various studies applied pair programming in programming labs and reported that it has more beneficial effects on performance than solo programming.

Synchronized programming (SP) has been used in computer science labs to enable students to code together at the same time in both distance and face-to-face classroom settings (Boyer, Dwight, Fondren, Vouk, & Lester, 2008; Schümmer & Lukosch, 2009). SP uses a plug-in with an IDE editor (Vandeventer & Barbour, 2012) to facilitate the program development pane for coding and compiling and the chat pane for online chatting. The SP system is a synchronized tasks system. When one user creates a task, that task generates an action transmitted to the other user. Synchronized actions consist of file operations (e.g., creating, saving, and removing files), editor operations (e.g., coding, code checking, and code integrating), and program running (i.e., execution and debugging).

**Virtualization-based lab**

Anisetti et al. (2007), Border (2007), and Wannous and Nakano (2010) introduced virtualization-based technology as a new way of installing a group of VMs on one server and running OSs on the VMs. Moreover, labs that implement virtualization-based technology allow learners to conduct experiments like real lab devices with flexible and portable features that have been successfully tested and verified for learning purposes.

Xu et al. (2014) have reported that there are few existing virtualization studies, and this research also found limited information about virtualization in existing studies, as shown in Table 1. We classified existing studies into four categories based on research variables; in reference to existing studies’ variables, this study’s subject is a course, and activities (i.e., individual, pair, or group tasks) are lab tasks assigned to students in its experiment. Table 1 demonstrates that the current study extends previous research, which focused only on students’ achievement and perception by including the instructor’s activity as a research variable and investigating how synchronized mechanisms of paired students can benefit learning of computer networks. Furthermore, this study conducted further comparison analysis, such as the t-test and ACOVA.
<table>
<thead>
<tr>
<th>Existing study</th>
<th>Variable</th>
<th>Subject</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ros, Robles-Gomez, Hernandez, Caminero and Pastor (2012)</td>
<td>Student’s perception</td>
<td>Networks configuration</td>
<td>Individual/Group Tasks</td>
</tr>
<tr>
<td>Anisetti et al. (2007)</td>
<td>Learning achievement</td>
<td>Computer Networks</td>
<td>Individual/Group Tasks</td>
</tr>
<tr>
<td>Chen and Tao (2012)</td>
<td>Learning achievement</td>
<td>Web security</td>
<td>Individual Tasks</td>
</tr>
<tr>
<td>Wannous and Nakano (2010)</td>
<td>Student’s perception</td>
<td>Computer networks</td>
<td>Individual Tasks</td>
</tr>
<tr>
<td></td>
<td>Student’s perception</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student’s perception</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xu et al. (2014)</td>
<td>Learning achievement</td>
<td>Computer networks</td>
<td>Individual/Group Tasks</td>
</tr>
<tr>
<td>Hwang et al. (2014)</td>
<td>Learning achievement</td>
<td>Computer networks</td>
<td>Individual/Group Tasks</td>
</tr>
<tr>
<td></td>
<td>Student’s activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The current study (SPC in VBLab)</td>
<td>Learning achievement</td>
<td>Computer networks</td>
<td>Individual/Pair/Group Tasks</td>
</tr>
<tr>
<td></td>
<td>Student’s activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student’s perception</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instructor’s activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *Group comparison with t-test. **Group comparison with analysis of covariance.

**Distributed cognition theory**

The distributed cognition concept explains interactions between people and technologies. Distributed cognition especially focuses on the study of teamwork, information transformation in the team, and problem solving by teammates (Flor & Hutchins, 1991; Hutchins, 1990).

*Figure 1. (a: Left side) the modified proposed distributed cognition system. (b: Right side) the original distributed cognition system (Flor & Hutchins, 1991, p. 41).*

In the 1990s, Hutchins conducted research in software engineering; his study introduced distributed cognition to explain collaborative activities and analyzed a pair of programmers performing a perfective software maintenance task by analyzing a complex cognitive system. This study reported that successful software development is viewed as a consequence not of any single programmer’s cognition but of an interaction of programmers and development artifacts in a system of distributed cognition (Flor & Hutchins, 1991).
Based on this theory, the current study designs a proposed distributed cognition system (see Figure 1a), in which the pair works only on their own monitor while they communicate with their partner through three ways: face-to-face, online discussion, or synchronized terminal. This protocol differs from the original distributed cognition system as shown in Figure 1b in which the pair works on two separate terminals using only face-to-face communication. In Figure 1, arrows represent the direction of communication. The bidirectional line between paired students is a face-to-face communication, the bidirectional line between students and terminals refers to students obtaining information from terminals and inputting data into terminals, and the bidirectional line between two terminals represents communication between paired students via chat window. The unidirectional line refers to paired students obtaining information from the partner’s terminal.

System framework

Based on the literature mentioned above, this study’s system framework in Figure 2 consists of three main elements: Pair users, SPC, and VBLab. SPC is a mechanism that requires a pair of users to work synchronously to complete their tasks. The VBLab is a networking infrastructure that can facilitate synchronized Web GUI for a pair of users to work together in the VM. The Web GUI and VMs components belong to VBLab. The Web GUI is a synchronized web-based interface developed based on the synchronized IDE editor (Vandeventer & Barbour, 2012). However, our synchronized web-based interface is more flexible than the original synchronized IDE editor because it can be used anywhere on the Internet without program installation and is easy to integrate with other web-based components. VMs are running OSs that operate as networking equipment. In summary, Figure 2 shows that users can work synchronously with their partners to complete work in VMs by following the concept of SPC with Web GUI.

Research questions

This study’s purpose was to assess how beneficial SPC and VBLab are for students’ learning and to answer the following research questions:

- When SPC is applied in VBLab, do students complete their assignments (homework and post-test objectives) better than those working solo in VBLab? Does SPC in VBLab reduce the instructor’s efforts to help students?
- What are students’ perceptions and behavioral intentions when using VBLab in the two groups?
- What are the differences in students’ behaviors and benefits from solo and synchronized pair mechanisms, as deduced from the interviews?

Method

Participants and procedures

The experiment was conducted during the summer semester (March–May 2014) at a university in Thailand. Participants consisted of 61 undergraduate students enrolled in two sections of a computer networks course. One section, with 32 students, was assigned as the control group, while the other section, with 29 students (13 pairs and 1 three-member group), was assigned as the experimental group.

Procedures of the experiment were based on four overall steps, as shown in Figure 3. The experiment was administered twice a week in three-hour increments. The same teacher lectured both groups with the same lab
topics (see Appendix 2), which consisted of two parts: Basic Labs (Labs 1–3) and Advanced Labs (Labs 4–6). In the lab class, both groups did their lab assignments using the VBLab. The experimental group conducted their lab assignments using the SPC mechanism, whereas the control group members worked individually.

**Learning activity designs**

In this experiment, learning activities consisted of Basic Labs assignments, Advanced Labs assignments, and homework. Before the lab started, experimental and control students were further divided into groups of four pairs and groups of four students, respectively. Details regarding the designed lab class and the homework assignments are as follows.

**Basic Labs assignments**

Basic Labs were individual assignments and consisted of Labs 1–3. Instructors prepared lab materials and assignments for both student groups. At the beginning of class, the teacher briefed students on the objective and contents of the experiment and gave them assignments to be completed by the end of the three-hour class period. Completed assignments were then presented to instructors for evaluation. In addition, students had both face-to-face and online synchronous discussion (OSD) with class members (e.g., group members, teaching assistants, and instructors).

**Advanced Labs assignments**

Advanced Labs were group assignments and comprised Labs 4–6. Students in both experimental and control groups were asked to collaborate with their fellow group members to complete lab assignments within the class period. Again, students had both face-to-face and OSD with class members. In Lab 4 (see Appendix 3), each student and each pair configured a file server, web server, database server, and FTP server. At the beginning of the assignment, one student or one pair configured one type of server and then explained how to do so to other group members. In Labs 5 and 6 (see Appendix 4), each student and each pair managed and configured one of the four routers. In Lab 5, students configured static routing to route four routers to communicate with each other. In Lab 6, students configured dynamic routing (i.e., routing information protocol: RIP) to route four routers to communicate with each other.

**Homework**

For control and experimental groups, the teacher prepared the same homework assignments, which consisted of a post-lab question aimed at improving students’ understanding of the lab contents. Students were allowed to use the VBLab to determine answers from the command search window and to redo assignments to confirm their answers.

**Research variables**

In this experiment, the following variables were defined and compared with each other as well as with overall learning achievement.

- Command count: the total number of Linux commands coded by a student using the VBLab for Labs 2–6
- The ratio of incorrect to total commands: ratio of incorrect to total Linux commands coded by a student using the VBLab for Labs 2–6
- Chat message count: total number of chat messages typed by a student using VBLab for Labs 2–6 relevant to the lab assignment
- Homework scores: homework scores for Labs 1–6
- Pre-test and post-test: pre-tests 1 and 2 are students’ exam scores before Basic Labs and Advanced Labs, respectively. Post-test 1 is the students’ midterm exam score, while Post-test 2 is the students’ final exam score.
- The number of face-to-face helps: total number of students’ face-to-face help meetings with instructors
The virtualization-based lab

This study used the VBLab to support lab activities, i.e., online networking lab with OSD and lab materials. The VBLab is a web-based lab (see Figure 4) and consists of four features that include the following aspects: (1) a web terminal that allows students to input command lines into the system for performing labs. This terminal synchronizes and shows input command lines from two students in the experimental condition and shows input command lines from only one student in the control condition. The web terminal is the key feature of this study. Control students input commands to the terminal individually, while the pair of experimental students help
partners enter commands in synchronized web terminals. When one of the pair (the driver) inputs a command, the partner (the navigator) observes its correctness. In addition, the VBLab with a synchronized web terminal allows the pair to swap driver and navigator roles immediately, as they require; (2) a chat feature that allows students to have OSDs for discussing and sharing Linux commands with each other because all lab assignments are text-based activities, especially for the long Linux commands. Therefore, OSD is a suitable tool for communicating with the class members; (3) a command search box that enables students to find the Linux command manual for providing guideline commands, which are collected from all assignments; and (4) for Labs 1–6, lab materials, which are lab sheets that consist of lab introductions, assignments, and homework. Students can download lab materials before lab class. The VBLab is a web-based platform that has a web terminal based on an open source Gate One web terminal emulator (McDougall, 2011), and the other three features (e.g., chat window and command search box) were developed by PHP and MySQL database. The four features are integrated by PHP and HTML coding and deployed to the web server, and students can access their services.

Results and discussion

These research results and pedagogical implications are presented in relation to each of the research questions, above, in the following three sections.

Learning and activity outcomes

Analysis of pre-test, post-test results, and homework scores

The pre-test aimed to ensure that both student groups had equivalent basic knowledge required for learning in the course. According to Table 2, the mean and standard deviation of pre-tests 1 and 2 of the experimental and control groups indicated significant differences in pre-test 1 ($t = −2.095, p = .041$) and pre-test 2 ($t = −2.119, p = .038$) for the two groups. However, pre-test 1 and 2 scores for both groups were lower than 3 (out of 10); thus, the two groups of students had the same level of abilities prior to taking the course.

Analysis of covariance (pre-test as covariate) was also used to investigate whether students in the two groups differed in their post-test performance when SPC was used or not used during a lab (see Tables 3 and 4). These analyses reveal a significant difference between students in the experimental and control groups: Basic Labs: $F (1, 58) = 7.421, p < .05$; Advanced Labs: $F (1, 58) = 36.403, p < .001$.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Experimental group Mean</th>
<th>Control group Mean</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test 1</td>
<td>1.69</td>
<td>2.13</td>
<td>−2.095</td>
<td>.041</td>
</tr>
<tr>
<td>Pre-test 2</td>
<td>2.17</td>
<td>2.69</td>
<td>−2.119</td>
<td>.038</td>
</tr>
<tr>
<td>Post-test 1</td>
<td>9.62</td>
<td>7.59</td>
<td>−2.095</td>
<td>.041</td>
</tr>
<tr>
<td>Post-test 2</td>
<td>10.00</td>
<td>5.47</td>
<td>−2.119</td>
<td>.038</td>
</tr>
</tbody>
</table>

Table 2. Results of pre-test and post-test analysis

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
<th>$F$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>73,822</td>
<td>1</td>
<td>73,822</td>
<td>7.421</td>
<td>.009**</td>
</tr>
<tr>
<td>Within group (errors)</td>
<td>576,978</td>
<td>58</td>
<td>9,948</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5120.000</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Analysis of covariance of post-test 1 performance with pre-test 1 as a covariate (Basic Labs)

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>$SS$</th>
<th>$df$</th>
<th>$MS$</th>
<th>$F$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>336,383</td>
<td>1</td>
<td>336,383</td>
<td>36.403</td>
<td>0.000***</td>
</tr>
<tr>
<td>Within group (errors)</td>
<td>535,944</td>
<td>58</td>
<td>9,240</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4417.000</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Analysis of covariance of post-test 2 performance with pre-test 2 as a covariate (Advanced Labs)

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

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

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

60
Moreover, t-test results of homework scores (see Table 5) show that the experimental group had significantly higher scores on Homework 1 ($t = 7.283; p = .00$), Homework 4 ($t = 2.485; p = .016$), and Homework 5 ($t = 9.742; p = .00$) than the control group; further, the average scores of Homework 2, 3, and 6 for the experimental group are higher than that of the control group. The primary reason for this difference is that experimental students usually discussed and answered homework questions with their partners and therefore attained higher homework scores.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Experimental group ($n = 29$)</th>
<th>Control group ($n = 32$)</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework 1</td>
<td>Mean 4.48, SD 0.91, SE 0.17</td>
<td>Mean 2.66, SD 1.04, SE 0.18</td>
<td>7.283</td>
<td>.00***</td>
</tr>
<tr>
<td>Homework 2</td>
<td>Mean 3.50, SD 0.82, SE 0.15</td>
<td>Mean 3.19, SD 0.68, SE 0.12</td>
<td>1.621</td>
<td>.11</td>
</tr>
<tr>
<td>Homework 3</td>
<td>Mean 4.21, SD 0.73, SE 0.13</td>
<td>Mean 4.16, SD 0.99, SE 0.17</td>
<td>0.226</td>
<td>.822</td>
</tr>
<tr>
<td>Homework 4</td>
<td>Mean 2.93, SD 0.65, SE 0.12</td>
<td>Mean 2.56, SD 0.50, SE 0.09</td>
<td>2.485</td>
<td>.016*</td>
</tr>
<tr>
<td>Homework 5</td>
<td>Mean 4.46, SD 0.87, SE 0.16</td>
<td>Mean 2.79, SD 0.35, SE 0.06</td>
<td>9.742</td>
<td>.00***</td>
</tr>
<tr>
<td>Homework 6</td>
<td>Mean 3.72, SD 0.80, SE 0.15</td>
<td>Mean 3.69, SD 0.78, SE 0.14</td>
<td>0.181</td>
<td>.857</td>
</tr>
</tbody>
</table>

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

Based on these findings, we can conclude that in both Basic and Advanced Labs, paired students in the experimental group performed better than solo students in the control group. We believe this is because they worked and synchronized with their partners in numerous discussions and engaged in knowledge-sharing to impart and receive more knowledge than might be gained through individual work (Lee, 2011).

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Experimental group ($n = 14$)</th>
<th>Control group ($n = 32$)</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab 2 Command count</td>
<td>Mean 103.79, SD 40.25, SE 10.76</td>
<td>Mean 53.84, SD 20.51, SE 3.63</td>
<td>4.399</td>
<td>.00***</td>
</tr>
<tr>
<td>Lab 3 Command count</td>
<td>Mean 49.14, SD 28.10, SE 7.51</td>
<td>Mean 41.34, SD 21.30, SE 3.76</td>
<td>1.035</td>
<td>.306</td>
</tr>
<tr>
<td>Lab 4 Command count</td>
<td>Mean 57.29, SD 23.84, SE 6.37</td>
<td>Mean 39.34, SD 23.22, SE 4.11</td>
<td>2.392</td>
<td>.021*</td>
</tr>
<tr>
<td>Lab 5 Command count</td>
<td>Mean 78.93, SD 34.60, SE 9.25</td>
<td>Mean 54.13, SD 21.54, SE 3.81</td>
<td>2.967</td>
<td>.005**</td>
</tr>
<tr>
<td>Lab 6 Command count</td>
<td>Mean 38.36, SD 14.01, SE 3.75</td>
<td>Mean 24.81, SD 16.82, SE 2.97</td>
<td>2.634</td>
<td>.012*</td>
</tr>
<tr>
<td>Lab 2 Ratio of incorrect to total commands</td>
<td>Mean 0.06, SD 0.03, SE 0.01</td>
<td>Mean 0.14, SD 0.10, SE 0.02</td>
<td>-4.142</td>
<td>.000***</td>
</tr>
<tr>
<td>Lab 3 Ratio of incorrect to total commands</td>
<td>Mean 0.07, SD 0.04, SE 0.01</td>
<td>Mean 0.16, SD 0.15, SE 0.03</td>
<td>-2.221</td>
<td>.032*</td>
</tr>
<tr>
<td>Lab 4 Ratio of incorrect to total commands</td>
<td>Mean 0.07, SD 0.04, SE 0.01</td>
<td>Mean 0.12, SD 0.07, SE 0.01</td>
<td>-2.823</td>
<td>.007**</td>
</tr>
<tr>
<td>Lab 5 Ratio of incorrect to total commands</td>
<td>Mean 0.05, SD 0.03, SE 0.01</td>
<td>Mean 0.08, SD 0.04, SE 0.01</td>
<td>-2.345</td>
<td>.024*</td>
</tr>
<tr>
<td>Lab 6 Ratio of incorrect to total commands</td>
<td>Mean 0.06, SD 0.05, SE 0.01</td>
<td>Mean 0.14, SD 0.10, SE 0.02</td>
<td>-3.148</td>
<td>.003**</td>
</tr>
</tbody>
</table>

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

**t-test results of command count and ratio of incorrect to total commands**

There are statistically significant differences regarding command counts in Lab 2 ($t = 4.339, p = .00$), Lab 4 ($t = 2.392, p = .021$), Lab 5 ($t = 2.967, p = .005$), and Lab 6 ($t = 2.634, p = .012$) between the experimental and control groups. Even though it was found that the command count was not directly related to learning achievement, this finding demonstrates that the command count is possibly an indicator affecting students’ learning. As indicated by Table 6, the mean of the lab command count of the experimental group was higher than the mean of the control group. This implies that the experimental group created more activities in command practices than the control group. Moreover, the significant differences in command count in Labs 2, 4, 5, and 6 show that paired students synchronized commands and helped each other perform assignments collaboratively in the VBLab. This led the experimental group to generate more commands than the control group (Braithwaite, Eby, & Wahls, 2008).
Furthermore, according to Table 6, ratios of incorrect to total commands show statistically significant differences between experimental and control groups in all six labs as follows: ($t = -4.142, p = .000$), ($t = -2.221, p = .032$), ($t = -2.823, p = .007$), ($t = -2.345, p = .024$), and ($t = -3.148, p = .003$). These results demonstrate that students in the experimental group improved their typing capability more than their counterparts in the control group because paired synchronization can facilitate careful and correct work; therefore, students in the experimental group could reduce syntax and typing errors (Lui & Chan, 2006).

**t-test results of chat message count and number of face-to-face helps**

Regarding *t*-test results of the chat message count, as shown in Table 7, there are statistically significant differences in the command count in Lab 4 ($t = -2.404, p = .019$), Lab 5 ($t = -2.501, p = .015$), and Lab 6 ($t = 2.737, p = .008$) between the experimental and control groups. Furthermore, almost all average values of the chat message count of the experimental group are lower than those of the control group, excluding that of Lab 2. In addition to the activity designs in the experiment, the students were asked to use OSD to communicate with class members for discussing, completing assignments, and requesting help. Therefore, the chat message count refers to students’ interactions with class members. According to the *t*-test results, compared with the control group, the experimental group shows lower interactions with class members. The main reason for this situation is that students in the experimental group could communicate directly with their partners to solve problems and to complete lab assignments efficiently; thus, compared with the control group that worked alone, the experimental group use less chat messages to accomplish assignments (Williams & Kessler, 2002).

Moreover, with regard to the number of face-to-face helps shown in Table 7, there are no significant differences between the experimental and control groups and the number of face-to-face helps are very limited. The cause for this situation could be the shyness of the Asian students when asking instructors for help (Liu, 2001). We now plan to deeply investigate the main reasons for the insignificant differences between the two groups in the number of face-to-face helps.

In conclusion, even though the two groups did the same assignments, the experimental group expended less effort than the control group in terms of chat interactions with group members, teaching assistants, and instructors.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Experimental group (n = 29)</th>
<th>Control group (n = 33)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>SE</td>
<td>Mean</td>
</tr>
<tr>
<td>Lab 2 Chat message count</td>
<td>7.00</td>
<td>5.56</td>
<td>1.03</td>
<td>6.93</td>
</tr>
<tr>
<td>Lab 3 Chat message count</td>
<td>2.34</td>
<td>2.09</td>
<td>0.39</td>
<td>3.72</td>
</tr>
<tr>
<td>Lab 4 Chat message count</td>
<td>4.14</td>
<td>3.82</td>
<td>0.71</td>
<td>6.53</td>
</tr>
<tr>
<td>Lab 5 Chat message count</td>
<td>2.24</td>
<td>2.54</td>
<td>0.47</td>
<td>4.31</td>
</tr>
<tr>
<td>Lab 6 Chat message count</td>
<td>0.72</td>
<td>0.88</td>
<td>0.16</td>
<td>1.28</td>
</tr>
<tr>
<td>Lab 1 Number of face-to-face helps</td>
<td>1.07</td>
<td>1.58</td>
<td>0.29</td>
<td>1.44</td>
</tr>
<tr>
<td>Lab 2 Number of face-to-face helps</td>
<td>1.00</td>
<td>1.10</td>
<td>0.20</td>
<td>1.78</td>
</tr>
<tr>
<td>Lab 3 Number of face-to-face helps</td>
<td>1.34</td>
<td>1.63</td>
<td>0.30</td>
<td>1.59</td>
</tr>
<tr>
<td>Lab 4 Number of face-to-face helps</td>
<td>1.59</td>
<td>1.70</td>
<td>0.32</td>
<td>2.09</td>
</tr>
<tr>
<td>Lab 5 Number of face-to-face helps</td>
<td>1.72</td>
<td>1.87</td>
<td>0.35</td>
<td>1.75</td>
</tr>
<tr>
<td>Lab 6 Number of face-to-face helps</td>
<td>1.28</td>
<td>1.87</td>
<td>0.35</td>
<td>1.31</td>
</tr>
</tbody>
</table>

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

**Students’ perceptions and behavioral intentions**

A questionnaire survey was conducted to investigate students’ perceptions and behavioral intentions when using the proposed system. The questionnaire was designed by following the technology acceptance model (Davis, 1986) for four dimensions (see questions numbered 6–9 in Table 1a in Appendix 1). It included five external
dimensions (see questions numbered 1–5 in Table 1a in Appendix 1) that affect intention to use and actual use (Davis, Bagozzi, & Warshaw, 1989). Responses obtained from the two groups were ranked using a five-point Likert scale (ranging from strongly disagree [1] to strongly agree [5]). Statistical results of the survey are presented in Table 1a in Appendix 1. According to the t-test results, there was a statistically significant difference regarding the perceived subjective norm by classmates to use the proposed system ($t = 2.036, p = .046$). This finding demonstrates that the subjective norm among students in the experimental group can enhance use and continued use of the proposed system in their future studies, as pairing helps them have more confidence and motivation to complete lab assignments (Williams, Wiebe, Yang, Ferzli, & Miller, 2002). Moreover, all questionnaire dimensions were rated “agree” except two dimensions of the control group (comprising perceived readiness for using the proposed system and behavioral intentions when using it) that were “neutral,” and all of the experimental group’s questionnaire ratings were higher than those of the control group. This indicates that the experimental group has more readiness and intention to participate in this learning style than the control group.

**Interviews and in-depth investigation**

During the one-on-one, semi-structured interviews, students mentioned that they could benefit from using SPC in VBLab for labs class. Four summaries of interviews are explained below.

Regarding the use of SPC for experiments, students in the experimental group indicated that SPC influenced their behavior—to have more discussions and share more—which related to the Linux commands and network configurations (Flor & Hutchins, 1991).

In addition, in SPC, students were typically able to find solutions to lower-level problems, such as syntax and typing errors, without assistance, and instances of asking instructors for help were much less frequent (Braught et al., 2011).

The use of SPC in our study was also intended to give students an opportunity to experience two different roles (driver and navigator) while completing the experiment. In either case, if one student in the pair was struggling or absent, the partner could swap roles and become the driver for the lab (Wiebe et al., 2003).

On the other hand, solo students in the control group frequently asked for help and had questions for instructors and other students both in face-to-face and online discussions of assignments (Wiebe et al., 2003). Furthermore, some students in the solo class had no self-confidence in doing assignments by themselves, and they did not ask others for help; thus, they did not finish assignments on time (Williams & Kessler, 2002).

**Implications regarding education and technology**

Based on these findings, the current study presents the following implications and recommendations for instructors who plan to teach computer networks. First, we recommend using SPC in VBLab for enhancing students’ understanding of lab contents. Secondly, SPC is a collaborative learning environment in which paired students can work and share knowledge for completing assignments; in addition, paired students and instructors can have direct conversations via a chat feature, and instructors can simultaneously monitor paired students in class and help them correct certain configurations by sending messages. Therefore, SPC can increase students’ attention when performing assignments during lab class. Thirdly, we observed that paired students had many online discussions with their partners during both Basic and Advanced Labs. Thus, instructors should chat directly into class chat window (not individual chat window) and discuss assignments with all pairs, especially when giving guideline commands. Fourthly, we recommend that instructors assign pairs based on students’ ability levels. Each pair should have at least one high-achieving student, because he/she will help his/her partner obtain better understanding of lab contents (Dawande, Johar, Kumar, & Mookerjee, 2008). Finally, this experiment shows that VBLab can be a cost-efficient lab solution when the alternative is to buy expensive high-profile network equipment directly from manufacturers.

**Conclusion**

This study applied the SPC mechanism for computer network labs to determine its effectiveness in students’ learning performance. In addition, the study investigated students’ perceptions and behavioral intentions when
implementing SPC in VBLab. First of all, this study successfully proposed SPC for learning in computer network labs. Second, the experimental group in the SPC class was more productive in command count; moreover, there was decreased effort from the instructor in terms of chat message count. Thus, the experimental group significantly outperformed the control group in both post-tests. Third, this study successfully deployed a VBLab into computer network labs, a technique that allowed instructors to create a variety of virtual network topologies to be deployed as lab facilities. Therefore, students learned with real experience and accepted VBLab for completing assignments during lab class. With regard to this study, one limitation is the relatively small sample used, which restricts broad generalization of results. Therefore, future studies should increase the number of participants. Finally, future efforts should compare performance of students working in pairs with synchronized and non-synchronized configurations.

References


### Table 8. t-test of questionnaire survey

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>Group</th>
<th>Dimension mean</th>
<th>SD</th>
<th>t</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>I think that the proposed system can provide a real Linux networking environment as working in a real machine.</td>
<td>Experimental</td>
<td>3.80</td>
<td>0.49</td>
<td>0.223</td>
<td>0.824</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.77</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>I think that the Virtual Box can provide a real Linux networking environment as working in a real machine.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>I think that the proposed system has good facilitation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4</td>
<td>I think that the proposed system has helpful peer and tutor support.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-1</td>
<td>I have no difficulty accessing and using Virtual Box.</td>
<td>Experimental</td>
<td>3.57</td>
<td>0.88</td>
<td>0.533</td>
<td>0.596</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.46</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-2</td>
<td>I have no difficulty accessing and using this system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>I think that I can remote to this system and it is stable everywhere.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>I think that I can access this system faster and smoothly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-1</td>
<td>I always peer review laboratory contents on the proposed system before class.</td>
<td>Experimental</td>
<td>3.43</td>
<td>0.60</td>
<td>1.746</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.16</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-2</td>
<td>I think that our educational (style) culture in class is ready for the proposed system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-3</td>
<td>I think that the proposed system makes students ready to do lab assignments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-1</td>
<td>I would like to collaborate with classmates in the same group for doing lab assignments.</td>
<td>Experimental</td>
<td>3.72</td>
<td>0.75</td>
<td>0.563</td>
<td>0.575</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.62</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-2</td>
<td>I would like to collaborate with classmates in another group for doing lab assignments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-3</td>
<td>I would like to share networks configuration and topology with group members for doing lab assignments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-4</td>
<td>From my experience, “collaboration” among classmates usually succeeds in finishing assignments faster.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-1</td>
<td>I think other students in my group should be aware of how to use this system.</td>
<td>Experimental</td>
<td>3.71</td>
<td>0.13</td>
<td>2.036</td>
<td>0.046*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.63</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-2</td>
<td>I think other students in my group would be willing to use this system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-3</td>
<td>I think other students in my classes would be willing to use this system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-4</td>
<td>Most people who are important to me think that it would be fine to use this system to do lab assignments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p < .05,* *p < .01,* **p < .001.
### Table 9. t-test of questionnaire survey

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>Group</th>
<th>Dimension</th>
<th>SD</th>
<th>t</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td><strong>Perceived ease of the proposed system use.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-1</td>
<td>I think that the proposed system is very convenient to do lab assignments.</td>
<td>Experimental</td>
<td>3.68</td>
<td>0.53</td>
<td>0.960</td>
<td>0.341</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.53</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-2</td>
<td>I think that the operation of the proposed system does not require too much time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-3</td>
<td>I think that the proposed system is very easy for doing practical lessons and exercises after class.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-4</td>
<td>I feel that learning to use this system is quite easy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-5</td>
<td>I think that the proposed system is very easy for communication with the instructor and other students.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td><strong>Perceived usefulness of the proposed system.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-1</td>
<td>I think that the chat window can communicate with other group members to have suggestions for accomplishing lab assignments.</td>
<td>Experimental</td>
<td>3.73</td>
<td>0.54</td>
<td>0.266</td>
<td>0.791</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.69</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-2</td>
<td>I think that the sharing of virtual network devices is helpful for doing lab assignments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-3</td>
<td>I think that the sharing of the chat window is useful for doing lab assignments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-4</td>
<td>I think that the proposed system increases collaborative work with other group members when doing lab assignments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-5</td>
<td>I think that the proposed system enhances my attention.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-6</td>
<td>I think that I can decrease my workload when working with the proposed system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-7</td>
<td>I think that I can memorize well when working with the proposed system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td><strong>Attitude toward using the proposed system.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-1</td>
<td>I like using this system to learn computer networks.</td>
<td>Experimental</td>
<td>3.78</td>
<td>0.69</td>
<td>1.367</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.56</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-2</td>
<td>I have a positive attitude toward using this system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-3</td>
<td>I feel that using this system to do lab assignments is a good method.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td><strong>Behavioral intentions when using the proposed system.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-1</td>
<td>If I have access to this system, I will use it to learn computer networks.</td>
<td>Experimental</td>
<td>3.56</td>
<td>0.68</td>
<td>1.204</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>3.36</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-2</td>
<td>If I do lab assignments, I will enjoy doing it with this system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-3</td>
<td>I think that I will use this system to help me when I do my homework.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, **p** < .01, ***p*** < .001.
Appendix 2. Topic of Laboratory assignment

Basic Labs: Linux concepts and basic practices
- Lab1 Introduction to Linux and Linux command
- Lab2 Linux Script
- Lab3 Install Linux

Advanced Labs: Linux networking and advance practices
- Lab4 Linux Networking, Configuration
- Lab5 Linux Networking, Static Routing
- Lab6 Linux Networking, Dynamic Routing

Appendix 3. Network topology of Linux networking, configuration: Lab4

Appendix 4. Network topology of networking, static routing and dynamic routing: Lab5 and Lab6
Technology-Enhanced Peer Review: Benefits and Implications of Providing Multiple Reviews

Pantelis M. Papadopoulos1, Thomas D. Lagkas2 and Stavros N. Demetriadiς3

1Aarhus University, Aarhus, Denmark // 2The University of Sheffield, International Faculty, CITY College, Thessaloniki, Greece // 3Aristotle University of Thessaloniki, Thessaloniki, Greece // pmpapad@tdm.au.dk // T.Lagkas@sheffield.ac.uk // sdemetri@csd.auth.gr

*Corresponding author

(Submitted January 12, 2016; Revised May 23, 2016; Accepted June 26, 2016)

ABSTRACT
This study analyses the impact of self and peer feedback in technology-enhanced peer review settings. The impact of receiving peer comments (“receiver” perspective) is compared to that of reaching own insights by reviewing others’ work (“giver” perspective). In this study, 38 sophomore students were randomly assigned in two conditions and engaged in peer review activity facilitated by a web-based learning environment asking them to provide multiple reviews. In the Peer Reviewed (PR) condition students both reviewed peer work and received peer comments for their own work. By contrast, in the Self Reviewed (SR) condition students provided peer reviews, but did not receive any. Instead, they were asked to perform self reviewing, before proceeding to any revisions of their work. Result showed that the two groups were comparable in all aspects, suggesting that the lack of getting peer reviews can be efficiently alleviated by other type of scaffolds such as a scripted self review process. Overall, the study provides evidence that the review “giver” perspective (as opposed to the typical “receiver” perspective) is a vital option and has noteworthy implications for the design of technological systems that aim to flexibly support more efficient peer review schemes.

Keywords
Peer review, Free selection, Technology-enhanced learning, Self review

Introduction

Peer review is a widely used instructional approach that has been proven valuable for assisting students in both the acquisition of domain-specific knowledge and the development of domain-independent skills. McConnell (2001) argued that peer reviewing offers students the opportunity for a constructive and collaborative learning experience, by engaging them in an active learning exercise. In its most common form, peer review entails four steps orchestrated by a teacher: (a) production of the initial student work, (b) assigning of reviewers, (c) feedback production, and (d) revisions. Each step poses additional research questions, while the literature abounds with variations of the method, each focusing on different learning goals.

Peer review can be implemented with or without technology, although the former enables the teacher in applying more complex instructional designs, by addressing larger audiences, supporting multiple reviews, and managing the process more efficiently. As the use of technology gets established in the typical classroom, the number of studies in the literature that use technology to support peer review activities increases steadily.

Our work is situated within the technology-supported peer review domain, focusing on settings where students play both the roles of reviewers and reviewees and provide multiple reviews, being also free to choose the peer work they want to review. We refer to this peer review method as “free selection.” Previous work showed that such freedom resulted into a greater amount of feedback produced, while increasing students’ level of engagement (Papadopoulos, Lagkas, & Demetriadiς, 2012). Several learning environments support multiple reviews, with varied levels of freedom allowed during the selection process (e.g., Denny, Hamer, Luxton-Reilly, & Purchase, 2008; Tsai & Liang, 2009; Tseng & Tsai, 2007; Cho & Schunn, 2007). Although researchers may control students’ free selection, an interesting research question that arises is what happens to the students that due to free selection do not receive any peer feedback. In other words, what are the benefits of providing peer reviews, and are they adequate to compensate for the loss of peer review comments?

This study provides new evidence on how student performance is affected by providing and receiving peer comments, thus informing both instructors and designers of technology-enhanced learning environments that support peer review.

In order to assist students that do not receive reviews, we asked them to perform self review, by explicitly reflecting on their answers and comparing them with others’. This approach was chosen because it is easy to
implement and it is based on students’ capacity and not on peers or the instructor. Regarding the effectiveness of the approach, articulating one’s own thinking and producing explicit external representations to reflect upon has been proven beneficial for learning (Jang, 2007; Nückles, Hübner, & Renkl, 2009). Although, self-reviewing could be performed by all students internally, making the process explicit is expected to increase critical reflection and support deeper understandings.

**Theoretical background**

**Peer review**

Peer review is a widely implemented didactic model, often supported by technological systems that lift the overhead imposed to the instructor. As the affordances of such systems are based on research evidence relevant to the value of various learning interactions that occur during the peer review activity, the current study focuses on providing additional evidence on the issue of giving vs. receiving feedback that is fundamental for the design of such systems.

Peer review is placed within the broader socio-cultural theoretical context (the “Vygotskian” approach) emphasizing the peer dialog and interaction as a key learning mechanism. Within the peer review model the “peer dialog/interaction” is typically implemented in an indirect asynchronous mode as peers are required to interact by reviewing each other’s work. The method is associated with higher-level learning skills, such as synthesis, analysis, and evaluation (Anderson & Krathwohl, 2001) as the students have the opportunity to analyse and evaluate peer work. Scardamalia and Bereiter (1994) have provided evidence that higher cognitive processes of learning are stimulated and guided by the peer review procedure, by implementing the method into school classes. The feedback provided through peer reviewing could be of greater quantity than the one provided by a busy instructor (Silva & Moreira, 2003; Wolfe, 2004), while the process of analysing peer work can support the development of students’ self-evaluation skills (Davies & Berrow, 1998), and improve their attitudes and self-efficacy (Anewalt, 2005).

The method has been used extensively in various fields such as second language writing (Hansen & Liu, 2005; Lundstrom & Baker, 2009; Rouhi & Azizian, 2013), writing instruction and relevant courses at the college level (Haswell, 2005), statistics (Goldin & Ashley, 2011), psychology (Cho & MacArthur, 2010), and computer science (Liou & Peng, 2009; Luxton-Reilly, 2009).

Researchers stress the fact that peer review offers to students the chance of developing a range of skills important in the development of language and writing ability, such as meaningful interaction with peers, a greater exposure to ideas, and new perspectives on the writing process (Hansen & Liu, 2005; Lundstrom & Baker, 2009). Certain studies support the use and adoption of peer review of writing (e.g., Cho & Schunn, 2007; Cho, Schunn, & Charnley, 2006), emphasising that when students get peer feedback and revise their written work, they improve their writing skills (Cho & MacArthur, 2010).

**Technology-supported peer review**

Peer review requires a significant volume of information exchange and may pose a difficult administrative overhead for the instructor. Technology can lift this overhead, by distributing material, collecting student work, granting access to peer work, guiding students in the review process, and providing a comprehensive picture to the instructor that orchestrates the process. Arguably, one of the greatest benefits of using technology in peer review is to have students perform multiple reviews. Addressing this issue, many studies used technology-enhanced learning environments and explored the benefits emerging from increasing the number of peer assessors (e.g., Tsai & Liang, 2009; Tseng & Tsai, 2007), by comparing single versus multiple peer reviews (Cho & Schunn, 2007). For example, Cho and Schunn (2007) reported that students that received feedback from multiple peers in the SWoRD web-based peer review system improved their writing quality more than students that received feedback from a single expert.

Systems such as Wolfe’s (Wolfe, 2004), PeerWise (Denny, Hamer, Luxton-Reilly, & Purchase, 2008), and curriculearn.dk (http://www.curriculearn.dk) support peer review settings without limiting the number of reviews a student can perform. In such systems, results showed that students with higher grades tend to contribute more than weaker students, resulting in a greater amount of higher quality feedback being produced (Luxton-Reilly, 2009).
The literature abounds with studies that use technology-enhanced learning environments around peer feedback. For example, Silva and Moreira (2003) supported peer interaction through the WebCoM system, while Liu and Tsai (2005) used web-based peer assessment to support conceptual awareness. Obviously, technology can enhance an already flexible and powerful instructional tool such as the peer review. However, the affordances of technological systems do have a significant impact on the efficiency of implementing any specific version of peer review technique. Thus, informed design approaches for improving learning systems to better support peer review-based learning are still in need.

**Giving vs. receiving feedback**

Although peer review is a well-known and in-depth explored learning technique, investigating the benefits emerging for peers from providing reviews – as opposed to simply receiving – is currently a critical issue in peer review literature. For example, Dunlap and Grabinger (2003) argued that reviewing someone else’s work can be beneficial for the student in reflecting on and articulating own views and ideas, thus eventually improving own work. In addition, Ko and Rossen (2004) stated that while reviewing others, students receive perspectives other than the instructor’s and this process could provide further insights.

Researchers refer to “assessors vs. assessees” (Li, Liu, & Steckelberg, 2010) or “givers vs. receivers” (Lundstrom & Baker, 2009; Rouhi & Azizian, 2013) and report different learning outcomes identified for these two different peer roles. In contrast to what is typically considered as the main peer review merit (that is, getting reviews), available research reveals that students who provide reviews to their peers (“assessors” or “givers”) reach higher levels of learning gains in comparison to students who typically only receive peer reviews (“assessees” or “receivers”). For example, Li et al. (2010) reported a significant relationship between the quality of students’ final projects and the quality of peer feedback the students provided, while no such relationship was identified for the feedback students received. Similarly, Lundstrom and Baker (2009) concluded that students who reviewed their peers’ writings were significantly benefited in their own writing, outperforming those students who only received peer feedback. However, studies exploring this issue are considered limited and authors suggested that further research is needed to explore the various scenarios and roles of students in assessor/assessee situations (e.g., Li et al., 2010). Finally, Rouhi and Azizian (2013) presented similar findings, with “givers” outperforming “receivers” in second language writing, emphasizing also that most studies have touched this issue rather superficially without providing strong empirical evidence.

**Research motivation, hypotheses, and research questions**

Based on the above background, our motivation is to deeper understand the impact of engaging students in different modes of peer-review activity. This, in turn, is expected to generate well-informed suggestions for the design of technological systems that support the activity. The current study focuses on the role that the two distinct processes play in peer review activity: (a) the externally originated review comments that guide students in revising their work (the “receiver” perspective), and (b) the insights developed by students themselves, based on the experience of offering reviews to their peers (the “giver” perspective).

In the study, we compared the performance of two groups. Although both groups provided reviews to peers, students in the Peer Reviewed (PR) group received reviews from fellow students, while students in the Self Reviewed (SR) group did not receive peer reviews and instead they were guided through a self review process. We tested the two following directional research hypotheses:

- H1(revision): “Students in the PR group will perform better in revising their own work than students in the SR group.”
- H2(conceptual): “Students in the PR group will perform better in acquiring domain conceptual knowledge than students in the SR group.”

In addition, the study also explored research questions regarding the attitudes and review strategies applied by the students and how these affected the learning outcome. More specifically:

- RQ1: How will the students select which peer work to review?
- RQ2: How will the volume and quality of peer work reviewed affect reviewer’s performance?
- RQ3: How will the volume and quality of peer comments received affect author’s performance?
- RQ4: How will the free selection peer process affect students’ attitudes towards the activity?
Method

Participants

The study employed 38 sophomore students (20-21 years old) who volunteered to participate. All students were majoring in Informatics and Telecommunications Engineering in a 5-year study program and were enrolled in the “Network Planning and Design” (NP&D) course. The activity was an optional part of the course, and a bonus grade was awarded to participants who successfully completed all phases. We randomly distributed students into the groups:
- Peer Reviewed: 18 students, 12 males and 6 females
- Self Reviewed: 20 students, 12 males and 8 females

Students were not aware of the distribution and we informed them beforehand of the research nature of the activity and the possibility that some of them would not receive peer reviews.

Learning environment

We have developed our own web-based learning environment that has been used in various domains in the past. The benefit of using a custom-made tool is that we could easily alter the study conditions according to our research interests. Studying in our learning environment typically entails answering open-ended questions of realistic scenarios, grounding the answers on related past cases that present similar problems and the way they were addressed. The learning environment implemented the free selection peer review protocol (described next), granting access to multiple peer work in the review phase.

Design

The study followed a pre-test post-test research design to compare the performance of the two groups. The study had six distinct phases: Pre-test, Study, Review, Revise, Post-test, and Interview. The study conditions were identical for all students throughout the activity, except from the Revise phase, in the beginning of which the PR group received the peer comments and the SR group performed self review. Two instructors of the NP&D course served as raters.

Procedure and study conditions

Pre-test phase

In the Pre-test phase, we recorded students’ prior domain knowledge in class using a written test with 6 open-ended questions (e.g., “How can the security requirements of a network affect its architecture?”).

Study phase

The Study phase started right after the pre-test and lasted one week. The students logged in the environment (from wherever and whenever they wanted) and worked on the available material, providing answers (in the following: solutions) to open-ended questions of 3 plausible scenarios (3 answers in total). Students had to take into account the specific conditions and the presented context and propose their own computer networks as solutions to the scenarios.

Review phase

Next, in the Review phase that lasted 4 days, the students had to review, in a double-blind process, the solutions their peers gave to the scenarios. All 38 participants had to provide reviews to the solutions of the 18 students of the PR group. In order not to overwhelm the students, for each of the 3 scenarios, participants received peer solutions from a randomly selected subgroup of 9 PR students. The solutions were presented in random order in a “solution grid” (Figure 1) and each grid was unique. However, each solution appeared the same number of times in total, thus having the same chances of getting a peer review.
Figure 1. The solution grid of a scenario. According to this figure, the student has read solutions 1, 2, 3, 5, 6, and 7, and has reviewed solutions 1 and 6 of this particular scenario.

Figure 2. Review form

The students were able to read as many peer solutions they wanted and they had to perform at least one review for each of the 3 scenarios. Of course, it was possible for some solution to receive more than one review, while for others to receive none. The latter, however, did not occur, since the number of reviewers (i.e., PR+SR = 38) was much higher than the number of solution authors (i.e., PR = 18), resulting to all solutions to receive at least one review.

Students had to follow a review microscript, guiding them through the process, focusing on three dimensions: (a) content, (b) argumentation, and (c) clarity (Figure 2). In the content section of the review, students were expected to analyse the main points of their peer’s solution, identifying for example the components of the proposed computer network. As students often fail to present adequately the reasoning behind their choices in proposing a
solution, the second guideline focused reviewers’ attention on argumentation. In other words, the reviewers had to evaluate whether the author justified each aspect of the network appropriately and if he/she grounded his/her solution on the provided material. The third guideline was focused on the clarity and eloquence of the solution, guiding reviewers into giving advice on the form of the solution. Finally, along with the comments, the reviewer had to also suggest a grade according to the following scale: (1: Rejected/Wrong answer; 2: Major revisions needed; 3: Minor revisions needed; 4: Acceptable answer; 5: Very good answer).

**Revise phase**

The Revise phase lasted 3 days. Right after the Review phase was completed, the produced feedback was made available to the students by the system. Students in the PR group received the comments their peers had made on their solutions, while students in the SR group did not receive any peer feedback. Instead, the SR group was reminded of the review form, and was prompted to perform self review (minor differences in wording existed between peer review and self review forms, to target one’s own solutions). As mentioned earlier, this was the only phase where the two groups worked in different ways. Once the system granted access to the peer comments, the PR group was able to proceed immediately to revise theirs solutions. On the contrary, students in the SR group had to submit self reviews, before they were allowed to revise their solutions.

**Post-test phase**

At the end of the online activity, the students took a written post-test in class focused on students’ acquired domain knowledge, using 3 open-ended questions (e.g., “Which network characteristics are affected by the end-users’ profile?”). The answers to these questions were not to be found as such in the study material, but rather to be constructed, by taking into account information presented in various cases.

**Interview phase**

Shortly after, students from each group were interviewed to record their approaches and comments on the activity. Interviews were semi-structured and focused on students’ views on the activity.

Figure 3 presents the phase sequence for each group (the study variables, marked in bold, are presented in the next section).

![Figure 3. Phase information and sequence](image-url)
Measures

The type of study conditions the students had before revising their initial solutions was the independent variable (i.e., peer reviewed or self reviewed), while students’ performance (in the written tests and in the learning environment) and activity (according to system log files) were the dependent variables. Table 1 presents each dependent variable related to students’ performance, along with the scale used and the way the total final score was calculated for each metric.

A 1-10 scale was used for the Pre- and the Post-test scores, while a 1-5 scale was used for the Solution.Rater, Solution.Peer, Reviewing, and Solution.Revised scores, to be in line with the scale used by the students in their review process.

Table 1. Dependent variables for students’ performance

<table>
<thead>
<tr>
<th>Score name (scale)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test (1-10)</td>
<td>The mean score the student received from the two raters for the 6 questions of the pre-test instrument.</td>
</tr>
<tr>
<td>Solution.Rater (1-5)</td>
<td>The mean score the student received from the two raters for the initial 3 solutions in the respective scenarios of the learning environment.</td>
</tr>
<tr>
<td>Solution.Peer (1-5)</td>
<td>The mean score the student received from peer reviewers for the initial 3 solutions of the learning environment.</td>
</tr>
<tr>
<td>Solution.Self (1-5)</td>
<td>The mean score the student assigned to his/her scenario solution in the self-review form.</td>
</tr>
<tr>
<td>Reviewing (1-5)</td>
<td>The mean score the student received from the two raters for the quality of each review he/she submitted.</td>
</tr>
<tr>
<td>Distance (0-4)</td>
<td>The mean absolute difference between the review scores submitted by the student and the respective scores submitted by the raters.</td>
</tr>
<tr>
<td>Solution.Revised (1-5)</td>
<td>The mean score the student received from the two raters for the revised 3 solutions in the respective scenarios.</td>
</tr>
<tr>
<td>Post-test score (1-10)</td>
<td>The mean score the students received by the two raters for the 3 questions of the post-test instrument.</td>
</tr>
</tbody>
</table>

In analysing students’ review strategies and the impact the review process had on their performance, the analysis included four additional dependent variables based on system log files that were related to students’ review strategy (Table 2).

Table 2. Dependent variables for students’ review strategy

<table>
<thead>
<tr>
<th>Variable name (scale)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NrReviewsSubmitted (3-27)¹</td>
<td>The number of reviews submitted by a student.</td>
</tr>
<tr>
<td>NrReviewsReceived (0-42)²</td>
<td>The number of reviews received by a PR student.</td>
</tr>
<tr>
<td>QltReviewsReceived (1-5)</td>
<td>The mean score of the quality of the reviews (as determined by raters) received by a PR student.</td>
</tr>
<tr>
<td>QltSolutionsReviewed (1-5)</td>
<td>The mean score of the quality of peer solutions (as determined by raters) reviewed by a student.</td>
</tr>
</tbody>
</table>

Note. ¹Each student sees a 3x3 solution grid in each of the 3 available scenarios and has to review at least one peer solution per scenario. ²Each PR student’s solution appeared approximately to 14 other students (PR and SR) and each student had to submit one question in the 3 available scenarios.

Data analysis

To avoid any biases, students’ paper sheets (pre- and post-test) and system print-outs (scenario solutions and reviews) were mixed and assessed blindly by the two raters, following predefined grading instructions. As a measure of inter-rater reliability, the two-way random average measures (absolute agreement) intraclass correlation coefficient (ICC) was calculated for the raters’ scores.

For all statistical analyses, a level of significance at .05 was chosen. Cronbach’s Alpha test was used to evaluate the internal consistency of the pre and post-test, while bivariate Pearson’s correlation test was used to calculate the respective correlation between the two test instruments. Bivariate Pearson’s correlation test was also used to analyse any interaction between the variables describing students’ performance (i.e., Table 1) and review strategies (i.e., Table 2).
Finally, each one-on-one interview lasted approximately 15 minutes and was audio recorded. The interview transcripts were used for content analysis.

Results

Student performance

Inter-rater reliability was high for the Pre-test (ICC = .882), the Solution.Rater (ICC = .849), the Reviewing (ICC = .892), the Solution.Revised (ICC = .880), and the Post-test (ICC = .843) scores. In addition, internal consistency was high for pre (α = .821) and post-test (α = .866), while the scores of the two instruments were not correlated (p > .05). Table 3 presents the results regarding the performance of the two groups throughout the activity.

Table 3. Student performance in the activity

<table>
<thead>
<tr>
<th>Written (scale: 1-10)</th>
<th>Peer reviewed</th>
<th>Self reviewed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>Pre-test</td>
<td>2.04 (1.06)</td>
<td>18</td>
<td>2.20 (1.09)</td>
</tr>
<tr>
<td>Post-test</td>
<td>8.13 (1.40)</td>
<td>18</td>
<td>8.19 (1.37)</td>
</tr>
<tr>
<td>Online (scale: 1-5)</td>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>Solution.Rater</td>
<td>2.73 (0.93)</td>
<td>18</td>
<td>2.43 (1.10)</td>
</tr>
<tr>
<td>Solution.Peer</td>
<td>3.32 (0.77)</td>
<td>18</td>
<td>--</td>
</tr>
<tr>
<td>Solution.Self</td>
<td>--</td>
<td>--</td>
<td>3.39 (0.90)</td>
</tr>
<tr>
<td>Reviewing</td>
<td>3.21 (0.78)</td>
<td>18</td>
<td>3.33 (0.59)</td>
</tr>
<tr>
<td>Distance (0-4)</td>
<td>1.06 (0.51)</td>
<td>18</td>
<td>0.98 (0.52)</td>
</tr>
<tr>
<td>Solution.Revised</td>
<td>3.28 (0.77)</td>
<td>18</td>
<td>3.29 (1.05)</td>
</tr>
</tbody>
</table>

T-test results did not indicate a significant difference between the two groups (p > .05) regarding their prior knowledge (Pre-test score), the quality of their initial solutions (Solution.Rater score), the quality of their submitted reviews (Reviewing score), the distance from raters’ score (Distance score), and the quality of their revised final solutions (Solution.Revised score), therefore suggesting that the two groups were comparable. One-way analysis of covariance also showed that there was no significant difference (p > .05) in the post-test performance (Post-test score) of the two groups.

In addition, paired-samples t-test results showed that in both groups the revised solutions were significantly improved, when compared to the initial ones (PR: t[17] = 4.618, p = .001, d = .66; SR: t[19] = 3.669, p = .003, d = .82), suggesting that the second week of the activity (Review and Revise phases) had an important impact on students’ learning.

Reviewing strategies and bivariate correlations

Usage data analysis showed that most of the students’ solutions in the scenarios were approximately half a page long, with the maximum length being one full page. Students read almost all the available solutions (M = 8.12, SD = 1.45) in the grid of each respective scenario and t-test results showed that the two groups were comparable (p > .05) regarding the number of reviews submitted and the quality of the solutions students chose to review (Table 4).

Table 4. System activity metrics

<table>
<thead>
<tr>
<th>Peer reviewed</th>
<th>Self reviewed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>NrReviewsSubmitted (3-27)</td>
<td>5.50 (0.92)</td>
<td>18</td>
</tr>
<tr>
<td>NrReviewsReceived (0-42)</td>
<td>10.77 (2.75)</td>
<td>18</td>
</tr>
<tr>
<td>QltReviewsReceived (1-5)</td>
<td>3.25 (0.51)</td>
<td>18</td>
</tr>
<tr>
<td>QltSolutionsReviewed (1-5)</td>
<td>2.33 (0.82)</td>
<td>18</td>
</tr>
</tbody>
</table>

Regarding the research questions of the student on students’ review strategy, Tables 5 and 6 present the bivariate Pearson’s correlation test results for the two groups. To increase readability, only the statistically significant correlations are presented. The two groups were comparable presenting the same bivariate correlations. This is in line with the previous results showing that students in the two conditions performed the same.

76
Table 5. Bivariate correlations in Peer Reviewed group

<table>
<thead>
<tr>
<th>Review strategy</th>
<th>Authoring skills</th>
<th>Reviewing skills</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-test</td>
</tr>
<tr>
<td>NrReviewsSubmitted</td>
<td>1</td>
<td>.503*</td>
</tr>
<tr>
<td>NrReviewsReceived</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>QltReviewsReceived</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>QltSolutionsReviewed</td>
<td>1</td>
<td>.577'</td>
</tr>
<tr>
<td>Pre-test</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Solution.Rater</td>
<td>1</td>
<td>.806**</td>
</tr>
<tr>
<td>Solution.Peer</td>
<td>1</td>
<td>.612'</td>
</tr>
<tr>
<td>Solution.Revised</td>
<td>1</td>
<td>.597'</td>
</tr>
<tr>
<td>Post_test</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Reviewing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .10; *p < .05; **p < .01.

Table 6. Bivariate correlations in Self Reviewed group

<table>
<thead>
<tr>
<th>Review strategy</th>
<th>Authoring skills</th>
<th>Reviewing skills</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-test</td>
</tr>
<tr>
<td>NrReviewsSubmitted</td>
<td>1</td>
<td>.534'</td>
</tr>
<tr>
<td>NrReviewsReceived</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>QltReviewsReceived</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>QltSolutionsReviewed</td>
<td>1</td>
<td>.604'</td>
</tr>
<tr>
<td>Pre-test</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Solution.Rater</td>
<td>1</td>
<td>.800**</td>
</tr>
<tr>
<td>Solution.Self</td>
<td>1</td>
<td>.612'</td>
</tr>
<tr>
<td>Solution.Revised</td>
<td>1</td>
<td>.559'</td>
</tr>
<tr>
<td>Post_test</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Reviewing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .10; *p < .05; **p < .01.

Table 7 combines and summarizes the most important findings (F#) of the bivariate correlation analysis.

Table 7. Findings of the bivariate correlation analysis

<table>
<thead>
<tr>
<th>#</th>
<th>Correlated variables</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>(Solution.Rater – QltSolutionsReviewed) and (Solution.Rater – NrReviewsSubmitted)</td>
<td>Students that have provided better initial solutions also submitted more reviews, selecting mainly high quality peer solutions to review.</td>
</tr>
<tr>
<td>F2</td>
<td>(Solution.Rater/Peer/Self – Reviewing – Solution.Revised)</td>
<td>Students that submitted good initial solutions (as determined by raters, peers, or themselves) also submitted high quality reviews and revised solutions.</td>
</tr>
</tbody>
</table>
F3 (QltSolutionsReviewed – Reviewing – Solution.Revised) Students that reviewed high quality peer solutions had also better performance in reviewing and revising.
F4 (Reviewing – Distance) Students that submitted high quality reviews suggested also grades closer to the ones suggested by the raters.
F5 (NrReviewsSubmitted – Reviewing) The higher the number of reviews submitted by a student, the higher the quality of the reviews. (Tentative)
F7 (Post-test -/- Reviewing) No correlation between post-test performance and reviewing.
F8 (NrReviewsReceived -/- any) and (QltReviewsReceived -/- any) The number and quality of received reviews were not correlated to any other variable in the study.

Interviews

The lack of peer comments in SR group did not raise any major concerns, with students in both groups evaluating positively their experience in the activity and the free selection peer review process. A small number of SR students (N = 6) said that they would like to receive peer comments, although they also expressed the belief that their revised solutions would not have been affected much by this.

Results showed that students submitted, on average, almost double the number of minimum reviews required (M = 5.71, SD = 0.88). Students mentioned the usefulness of forming and articulating a written review in clarifying their own understandings, while the format and length of the solutions in the activity made the review process easier.

Students in PR group found the peer comments generally useful claiming that they decided to revise their solutions after a comment they received. Of course, there were also students that expressed disagreement with reviewers’ opinions. Similarly, students in SR group stated that the self review helped them to better analyse their initial solutions, mentioning that the structured process of looking back to their own work after having reviewed others’ made their weakness more visible to them. However, many students said that they had taken the decision to revise their solutions even before self review, usually because of something they had read in peers’ solutions. More specifically, students mentioned that there were cases in which peer solutions presented arguments which “made more sense” than theirs or “presented the same solution to the scenario, but in a more clear way.”

Discussion

Hypotheses testing

The comparison between the two conditions showed that students in the two conditions performed the same in every way. In the light of these results, research hypotheses H1(revision) and H2(conceptual) are rejected, suggesting that the process of giving and receiving peer reviews was not enough for the Peer Reviewed group to outperform the Self Reviewed group that gave peer reviews but did not receive any, performing self review instead. As such, alternative hypotheses are stated:

- Ha1(revision): “Students in the PR group will not perform better in revising their own work than students in the SR group.”
- Ha2(conceptual): “Students in the PR group will not perform better in acquiring domain conceptual knowledge than students in the SR group.”

Research questions analysis

Analysing RQ1, finding F1 of the bivariate correlation analysis (Table 7) suggests that students’ selection strategy was correlated to their own levels of understanding, meaning that strong students tend to review more solutions, selecting also higher quality peer work. On the contrary, less competent students selected peer
solutions at their own levels of understanding (weaker solutions), while analysis shows that they reviewed lower numbers of peer solutions, thus indicating minimum effort strategy.

Regarding authoring, reviewing, and revising skills, finding F2 suggests that authoring an initial solution, reviewing peer work, and revising own work are all correlated. Although correlation analysis does not refer to causality, having in mind the sequence of the study phases, it seems that the level of knowledge students acquire during the first week of the activity is indicative of their performance in the phases to come. In other words, strong students are able to perform well in all the phases, while weak students struggle throughout the activity. This is important, because it dictates the need for an additional instructional intervention that would improve the performance of weak students. Such an intervention could be based on finding F3 and students’ interview comments in which they appreciated clear and well-grounded peer solutions. So, addressing RQ2, our assumption is that students may benefit more when they spend their time reviewing good solutions. We maintain that the deeper insights and better grounded argumentation found in higher quality peer work underlines the discrepancy between strong and weak solutions, and supports the generation of such self-constructed deeper understanding that could trigger weak students to improve their work. Having in mind also F1, it appears that weak students tend to select lower quality solutions, thus missing the opportunity to get deeper insights while reviewing. So, returning to the question raised earlier, one beneficial intervention could be the promotion of high quality peer work for reviewing, in an effort to guide weak students to a more useful review process. In our study, for example, this could be done by selecting only high quality solutions in the grids presented to weak students. Of course, this approach needs further validation through future research.

Further in the analysis, findings F4 and F5 were expected. Students that, according to the raters, provide better feedback to their peers are expected to also grade their peers similarly to raters, while the students’ ability to review peer work is expected to improve as students spend more time reviewing.

Finally, as far as students’ performance in acquiring domain knowledge is concerned, findings F6 and F7 showed that the scores achieved in the post-test were strongly correlated to the quality of the initial solutions, the quality of the revised solutions, and the quality of solutions reviewed, but there was no correlation between post-test scores and students’ skills in reviewing peer work. One explanation for this could be that the post-test instrument was strongly focused on conceptual domain knowledge, and as such it might not be suitable to also assess students’ reviewing skills. This does not mean that the review process itself did not have an impact on learning outcomes. Results already showed existing correlations between Reviewing score and other dependent variables, and a significant improvement between initial and revised solutions. On the other hand, addressing RQ3, F8 suggests that neither the number nor the quality of peer comments received have an effect on students’ performance.

Regarding RQ4, results showed that the majority of students submitted almost double the number of required reviews. In explaining this attitude, students cited (a) the usefulness of the review process for their own benefit, and (b) the short length of the peer solutions. Students’ ability to link the process of formulating and writing an analysis on someone else’s work with clarifying their own understandings suggests metacognition. However, a research question is raised here regarding the efficiency of the same review setting for students with lower metacognitive skills. Such question could be the focus of future research on peer review, as the level of metacognition could vary and affect students’ engagement in a learning activity significantly. Students’ positive attitude towards the review process applied is not new, as it has also appeared in a previous work of ours on defining and comparing the free selection protocol with assigned review settings (Papadopoulos et al., 2012). Furthermore, the students’ attitudes are in line with studies that report students’ tendency to contribute more when not restricted by the assignment protocol (e.g., Denny et al., 2008; Luxton-Reilly, 2009; Wolfe, 2004). Nevertheless, since all the students in this study volunteered to participate, one should also consider that students’ engagement may differ in settings where such activity is obligatory. Finally, students’ claim that in some cases the decision for revising their solutions was taken while reading peer solutions, demonstrates the impact the process of providing reviews could have on students’ performance.

A possible explanation

In peer review activity, peers are generally expected to deeper understand the domain and improve their contributions based principally on feedback comments offered by other peers. However, in a context where peers are offering multiple reviews they may also get significant benefits from the process of reviewing others’ work. The current study provided evidence that the reviews offered by other peers did not help the PR group students to outperform those in SR group who did not receive this kind of comments. Thus, reviews from peers may not
be of substantial help when compared to the insights that students develop, when reviewing others’ work and are prompted explicitly to self review their work. We suggest that there are various reasons leading to failure of peer review comments to further improve the PR performance: (a) PR students may ignore peer reviews, doubting their quality, as they do not originate from an expert, (b) poor quality of peer review does not actually help PR students to improve their solutions, and (c) peer reviews are of high quality, but do not actually offer any additional insights for improvement that the PR students could not develop themselves while reviewing others.

Additionally, as Lundstrom and Baker (2009) suggested, review providing is a cognitive activity that can be understood within the framework of socio-cognitive perspective (Vygotsky, 1986) relevant to student-provider’s zone of proximal development (ZPD). It seems that reviewing peers’ contributions acts as a scaffold that enables student-providers to successfully build understandings within the limits of their own ZPD, thus achieving to improve their performance as authors and reviewers. It might be also possible that this cognitive mechanism is further facilitated by the positive influence of affective and motivational factors, as student-providers develop ownership of their newly constructed understanding and are more willing to accept the corrective feedback implied. This, of course, is a claim open to future research and validation.

**Implications**

Free selection had positive effects on students’ attitudes towards the activity. At the same time, it is clear that not all selection strategies work the same for the students. Based on correlation analysis and students’ statements, we maintain that there is a possible connection between reviewing high quality peer work and higher student performance. A design implication would be to promote this high quality peer work for review, especially to weak students, perhaps by allowing students to rate the quality of peer work and designing the learning system in a way that encourages students to review high level work. The way this approach can be implemented without lessening the role of free selection depends on the context of the learning activity.

**Conclusions**

This study provides concrete evidence showing that in free selection peer review settings, where conducting multiple reviews is possible, the lack of peer reviews can be easily alleviated by supporting students through a simple self review process. We are not suggesting, of course, the elimination of peer comments. Students also expressed positive opinions for some of the peer reviews they received. The question that rises naturally is how the students are going to make use of free selection and how the spread of reviews will cover all the peer work. The instructor can opt to modify the setting in an effort to ascertain that each student will receive peer reviews. However, what the findings of this study suggest is that such a modification may not be necessary. Thus, systems for peer review can be designed to offer “receiving peer reviews” and “practicing self review” as equally important alternatives for the instructor to choose from.

**References**


Effects of a Structured Resource-based Web Issue-Quest Approach on Students’ Learning Performances in Computer Programming Courses

Ting-Chia Hsu1 and Gwo-Jen Hwang2*

1Department of Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // ckhsu@ntnu.edu.tw // gjhwang.academic@gmail.com

*Corresponding author

(Submitted January 14, 2016; Revised April 21, 2016 Accepted May 14, 2016)

ABSTRACT

Programming concepts are important and challenging to novices who are beginning to study computer programming skills. In addition to the textbook content, students usually learn the concepts of programming from the web; however, it could be difficult for novice learners to effectively derive helpful information from such non-structured open resources. Few studies have addressed this issue by proposing different web issue-quest approaches and investigating the effects of learning with different web resources on students’ learning performances in computer programming courses, not to mention the use of statistical methods for analyzing the factors affecting students’ learning outcomes. Therefore, in this study, a structured-resource issue-quest approach is proposed to support structured programming learning activities. An issue-quest learning environment was developed, and an experiment was conducted to compare the learning achievement of the students who learned with the proposed approach (experimental group) with that of the students who learned with the conventional open-resource-based issue quest approach (control group). The participants were 418 tenth graders (214 in the experimental group and 204 in the control group) who were 16 years old on average. Structural Equation Modeling (SEM) was employed to investigate and compare the effects of the Web Issue-Quest approaches (i.e., open resource-based issue quest and structured resource-based issue quest) on the students’ learning perceptions based on an extended technology acceptance model (ETAM). The experimental results showed that the structured resource-based issue quest approach benefited the novices more in terms of their learning achievements related to the computer programming concepts. Meanwhile, the SEM analysis showed that the two groups of students had equivalent perceptions of the usefulness, ease-of-use and enjoyment of learning with the two issue quest modes. This implies that the structured resource-based issue quest approach was helpful to the students in terms of improving their learning performance, while the limitation of access to open resources did not affect their learning perceptions.

Keywords

Computer programming, Issue quest, Structured resource, Technology Acceptance Model, Structural Equation Modelling

Introduction

Computer knowledge and programming skills have been recognized as a core competence in the 21st century (Kirby & Riley, 2006). In the past decade, educational institutes all around the world have considered basic computer knowledge and programming skills as a fundamental curriculum at all school levels (Esteves, Fonseca, Morgado, & Martins, 2011; Jeon, Kim, Hong, & Kim, 2014; Kordaki, 2010). For example, in the United States, the issue of teaching computer programming languages in school has been widely discussed for decades (Soloway, 1993). In Taiwan, students learn basic computer knowledge since the third grade, and basic programming skills since tenth grade; moreover, many colleges consider computer programming language courses as a fundamental curriculum for all graduate students (Wang, Huang, & Hwang, 2015). In Thailand, the higher education sector of the Ministry of Education has promoted computer and information technology skills, including computer programming concepts and skills in colleges since 2000 (Chookaew, Wanichsan, Hwang, & Panjaburee, 2016).

Meanwhile, programming has been identified as a course with a high drop-out rate (Gomes, Areias, Henriques, & Mendes, 2008). Researchers have pointed out several factors affecting students’ programming skills, including the lack of practice (Chen, Chang, & Wang, 2008) and insufficient or incorrect fundamental concepts of programming (Eckerdall, 2009). For example, students who lack the concept of the structure of programming are likely to write ill-structured programs (Ala-Mutka, 2004). In addition to their textbooks, most novices acquire computer and programming knowledge from the web, which provides non-structured open resources comprising various kinds of information. Therefore, many studies have reported the importance of conducting web-based learning activities for strengthening students’ competences of acquiring knowledge (Kuo, Hwang, & Lee, 2012; Land & Greene, 2000; Rae & Samuels, 2011; Taradi, Taradi, Radić, & Pokrajac, 2005). The aim of facilitating
student-centered learning through web technologies is to foster students’ abilities of data collection, extraction and application, as well as dealing with forthcoming challenges or problems (Ates & Cataloglu, 2007; Pimta, Tayrakham, & Nuangchalerm, 2009; Land & Greene, 2000). Researchers have called such web-based learning activities that engage students in searching for information on the web, selecting information, abstracting important and relevant content and summarizing their findings for investigating a specified issue or topic “Web Issue-Quest” (Sung, Hwang, & Chang, 2015).

Several previous studies have reported that students’ Web Issue-Quest ability could be insufficient if they lack experience or are not well trained (Bilal, 2001; Bilal, 2002). Researchers have further indicated that the type of web resources could also be another important factor affecting students’ learning perceptions and outcomes in Web Issue-Quest learning activities (Jonassen, 1991; Merrill, 1991); that is, in such an activity, the information sources play an important role. Students’ learning perceptions and performances could be significantly affected by the structure and content of the information sources (Fessakis, Gouli, & Mavroudi, 2012; Taradi, Taradi, Radić, & Pokrajac, 2005; Tsai, Lee, & Shen, 2013).

However, few studies have been conducted to investigate the effects of learning with different web resources on students’ learning performances in computer programming courses, not to mention the investigation of factors affecting their learning outcomes using statistical analysis (Jaeger & Adair, 2014; Senocak, 2009). Therefore, in this study, a structured resource-based web issue-quest approach is proposed to facilitate programming concept learning. An experiment has been conducted to evaluate the effect of the proposed approach on students’ learning achievements and perceptions in the “structured programming” unit of a high school fundamental computer programming course.

**Literature review**

The importance and strategies of teaching computer programming in schools have been widely discussed by scholars around the globe for decades (Robins, Rountree, & Rountree, 2003; Yang, Hwang, Yang, & Hwang, 2015). For example, Al-Bow et al. (2009) reported a game development strategy for teaching computer programming to high school students and teachers. Wang et al. (2015) conducted a project-based computer programming activity in a high school and found that mathematics-gifted students tended to have better learning outcomes than students who were not gifted in this area.

To further investigate the status of computer programming education around the globe, Lahtinen, Ala-Mutka and Järvinen (2005) conducted an international survey. They found that the Internet had become an important resource for learning computer programming. Consequently, scholars have suggested that the learning environment for programming should not be restricted to schools, but that online resources should also be incorporated into the programming curriculum (Gomes, Areias, Henriques, & Mendes, 2008). Among the various web-based learning approaches, Web Issue-Quest is widely adopted in school settings. It is a student-centered learning activity in which students search for new knowledge and learn subjects through a series of information seeking, selecting, abstracting and summarizing activities that eventually help them answer questions and comprehend the issue in depth (Kuo, Hwang, & Lee, 2012; Smith & Hung, 2017). The questions, ranging from easy to difficult, guide the learners through the process of gathering information, integration, and argumentation (Bradley et al., 2008; Golanics & Nussbaum, 2008; Oh & Jonassen, 2007).

To deal with a web-issue quest, students require the abilities of retrieving, recognizing and synthesizing information; moreover, they need to be capable of generating solutions based on the information they collect and organize (Brand-Gruwel, Wopereis, & Vermetten, 2005; Eisenberg, Johnson, & Berkowitz, 2010). In the meantime, by engaging students in such web-quest learning activities, teachers can observe how the students inquire about issues according to their existing knowledge. Via Web Issue-Quest learning activities, students are active in seeking information, collecting data, making selections, linking new and old knowledge, as well as developing and inferring based on their findings in the learning activity (Fleissner, Chan, Yuen, & Ng, 2006; Chan, 2007; Jwaifell & Al-Atyat, 2015). Researchers have indicated that such well-structured and guided inquiry activities are helpful to students in terms of developing their critical thinking ability (Choi, Lindquist, & Song, 2014; Kong, Qin, Zhou, Mou, & Gao, 2014; Martyn, Terwijin, Kek, & Huijsjer, 2014; Sommers, 2014).

Researchers have also pointed out that the provision of supportive resources for online courses could affect students’ perceptions of web-based learning activities (Palmer & Holt, 2008; Ucar & Trundle, 2011). Jwaifell and Al-Atyat (2015) indicated that providing quality information sources is one of the important components in web-quest learning activities for promoting students’ learning motivation and inquiry ability. Gordon and
Brayshaw (2008) further reported that web issue quest activities were helpful to students in terms of improving their learning achievements based on the experimental results in a university. Accordingly, such learning activities have great potential for contributing to students’ acquisition of knowledge (Şendağ & Odabaşı, 2009; Spronken-Smith, Bullard, Ray, Roberts, & Keiffer, 2008).

As few studies related to information seeking or Web Issue-Quest have focused on the effect of such an approach on students’ programming knowledge, or have compared the learning performance and perceptions of the students who engaged in web issue-quest activities using different information resources, this study aimed to address these issues in the “structured programming” unit of a high school computer course. Moreover, the TAM (Technology Acceptance Model) was adopted to evaluate the students’ learning perceptions. The TAM, proposed by Davis (1986), is a widely adopted model for investigating users’ perceptions of information systems. It provides a theoretical foundation that allows researchers to examine how external variables (e.g., the quality of the search environment or the user’s perceived enjoyment) affect users’ inner cognitions and attitudes (e.g., perceived ease of use and usefulness), thus resulting in behavioral intention to use information technologies (Davis, 1989; Venkatesh & Davis, 1996). Liaw and Huang (2003) further proposed an extended TAM Model by considering individual computer experience and the quality of the search systems. In this study, the extended TAM model of Liaw and Huang (2003) was adopted to more precisely investigate the perceptions of the students who learned with different web issue-quest resources.

**Structured and open resource-based issue quest systems**

In this study, an open resource-based and a structured resource-based issue quest system were employed to support the web-quest activities. The two systems have similar functions and interfaces for both teachers and students. The only difference between them is the information source, as shown in Figure 1. In the open resource-based system, students search for information via a metasearch engine, which invokes the existing search engines, such as Google, to search for information on the Internet (Meng, Yu, & Liu, 2002). On the other hand, the structured resource-based system searches for information from a database of an educational system. The database contains more than 500,000 data items, including 75% science-related materials, around 20% social science-related materials, and around 5% related to computer science; in particular, approximately 3% of the resources are related to the learning content of computer programming.

![Diagram of web issue-quest learning system](image)

*Figure 1. Structure of the web issue-quest learning system*

Figure 2 shows the interface of the issue-quest learning system. The “question statement” area depicts the questions proposed by the teachers, while the “answer” area allows them to submit their answers to the questions.
after invoking the “search” function to search for relevant information, select searched data, abstract relevant content, and summarize their findings. During the learning activity, the students first read the questions related to the quest issue on the left of the browser. They then fill in keywords to search for information for answering the questions. Following that, they need to select the searched results to browse, abstract relevant content to the answer area, modify the statements, and then submit their answers.

Research questions

In this study, we attempted to examine the learning achievement and perceptions of the students who learned with the structured resource-based and open resource-based issue quest approaches. Therefore, the following research questions were investigated:

- Do the students who learn with the structured resource-based issue quest approach show better learning achievements than those who learn with the conventional open resource-based issue quest?
- What are the perceptions of the students who learn with the structured resource-based issue quest from the perspective of the extended TAM model?
- What are the perceptions of the students who learn with the open resource-based issue quest from the perspective of the extended TAM model?

Method

In this study, the course unit “basic concepts of structured programming” in a vocational high school was adopted for conducting the experiment. The objective of the unit was to foster the concepts of three basic programming structures (i.e., serial, selective, and repeated structure) and their application criteria.

Participants

The subjects included twelve classes of tenth graders of a vocational high school in Taiwan. A total of 418 students, whose average age was 16, participated in the study. Six classes, totaling 214 students, were assigned to be the experimental group, while the other classes consisting of 204 students were the control group. The students in the experimental group learned with the structured resource-based issue quest approach, while the control group students learned with the open resource-based issue quest approach. There was no significant difference between the prior knowledge of the two groups because a pre-test had been conducted to confirm their related knowledge for this subject before the experiment. In addition, the IQ and background knowledge of the participants were seen as having no significant difference because they had just passed the entrance examination and had been assigned to the same school based on their performance in the examination.
Experimental procedure

Figure 3 shows the experimental procedure of this study. Before the learning activity, the students received 100 minutes of instruction on the basic concepts of the computer programming language. They then took the pre-test to evaluate their prior knowledge of learning structured programming. Following that, the teacher introduced the functions and operational procedure of the Web Issue-Quest systems and asked the students to practice operating the systems.

The web issue-quest activity lasted 100 minutes. During the learning activity, the students in the experimental group and the control group were engaged in answering the four questions using the open resource-based and the structured resource-based Web Issue-Quest systems, respectively.

After the learning activity, all of the students took a post-test and completed questionnaires on the perceived quality of the Web Issue-Quest systems as well as their enjoyment, ease of use, usefulness, and user intention regarding the learning activity.

Measuring tools

The pre-test and post-test were developed by an experienced high school teacher who had taught the computer programming course for more than ten years. The aim of the tests was to examine the students’ structured programming concepts. Each of the tests consisted of 15 multiple-choice and five short-answer questions with perfect scores of 100.
To assess the students’ learning perceptions, a modified version of the extended Technology Acceptance Model proposed by Liaw and Huang (2003; 2006) was adopted as the measuring tool. A total of 15 items in the questionnaire were used to assess the five latent variables of the causal model, including the quality of the Web Issue-Quest system (QOS), perceived enjoyment (PE), perceived ease-of-use (PEU), perceived usefulness (PU), and user intention (UI), as shown in Table 1. A 7-point Likert scale was used in the questionnaire, where 1 represented “Strongly Disagree” and 7 represented “Strongly Agree.” The Cronbach’s α values were 0.95, 0.92, 0.88, 0.95 and 0.91 for system quality (QOS), enjoyment (PE), ease of use (PEU), usefulness (PU) and intention (UI), respectively.

<table>
<thead>
<tr>
<th>Latent variables</th>
<th>Questionnaire items used as observed variables</th>
</tr>
</thead>
</table>
| Quality of the Web Issue-Quest learning system (QOS) | QOS1: I am satisfied with the information searching and resources.  
QOS2: I am satisfied with the quality of information generated from the information searching and resources.  
QOS3: I am satisfied with the functions of the information searching and resources. |
| Perceived Enjoyment (PE) | PE1: I am satisfied with using the information searching and resources to find online information.  
PE2: I like to use the information searching and resources to find information.  
PE3: I enjoy using the information searching and resources when I need to use them. |
| Perceived Ease of Use (PEU) | PEU1: It is easy to remember how to perform tasks using the information searching and resources.  
PEU2: My interaction with the information searching and resources is clear and understandable.  
PEU3: Overall, I find the information searching and resources easy to use. |
| Perceived Usefulness (PU) | PU1: Using the information searching and resources enhances my learning effectiveness.  
PU2: Using the information searching and resources makes learning more efficient.  
PU3: Overall, I find the information searching and resources helpful to me in acquiring knowledge. |
| User Intention (UI) | UI1: I believe that the information searching and resources are effective for finding information.  
UI2: I will use the information searching and resources to find information.  
UI3: I intend to use the information searching and resources to find information in the future. |

**Table 1.** Five latent variables and their observed items

**Question design for the Web Issue-Quest activities**

Previous studies have reported the question design principles for Web Issue-Quest activities (Hwang, Tsai, Tsai, & Tseng, 2008; Hsu, Hwang, Chuang, & Chang, 2012). Usually four questions are needed to guide students to investigate a specified issue via web information searching in a systematic way. The first, second, and third questions are structured problems for evaluating the students’ information searching (Keyword-adopting and information-selecting), abstracting, and organizing performance (Connell & Abramovich, 2017). The fourth is an open question for engaging students in critical and creative thinking (Reasoning and elaborating) based on what they discovered from answering the other questions (Kuo, Hwang, Chen, & Chen, 2012; Kuo, Hwang, & Lee, 2012). Table 2 illustrates the four questions designed for this study. The students’ cognition and concept of structured programming can be developed by dealing with the four Web Issue-Quest questions.

<table>
<thead>
<tr>
<th>1. Which basic structures are there in structured programming?</th>
<th>Keyword-adopting and information-selecting</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. What are the basic procedures of those structures in structured programming?</td>
<td>Information-abstracting</td>
</tr>
<tr>
<td>3. What are the advantages and disadvantages of structured programming?</td>
<td>Information-organizing</td>
</tr>
<tr>
<td>4. Do you agree or disagree with utilizing structured programming? Why?</td>
<td>Reasoning and elaborating</td>
</tr>
</tbody>
</table>

**Table 2.** Four questions for guiding students to investigate the concept of structured programming
Results

Analysis of learning achievement

One of the objectives of this study was to examine the effectiveness of the proposed approach in terms of improving the learning achievement of the students. ANCOVA was used to exclude the difference between the prior knowledge of the two groups by using the pre-test scores as the covariate and the post-test scores as dependent variables. The homogeneity test result showed that the post-test scores of the two groups were homogeneous \((F = 0.25, p = .62 > 0.05)\), implying that ANCOVA could be applied.

Table 3 shows the ANCOVA of the posttest scores of the two groups by excluding the impact of their pretest scores. It was found that the students in the experimental group had a higher adjusted mean (i.e., 13.84) than those in the control group with \(F = 11.06\) and \(p < .05\), showing that learning with the structured resource-based issue quest activity benefited the high school students more than learning with the open resource-based issue quest approach.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjust mean</th>
<th>(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>204</td>
<td>12.65</td>
<td>4.78</td>
<td>12.79</td>
<td>11.06*</td>
</tr>
<tr>
<td>Experimental group</td>
<td>214</td>
<td>13.94</td>
<td>4.07</td>
<td>13.84</td>
<td></td>
</tr>
</tbody>
</table>

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

Analysis of learning perceptions

Structural Equation Modeling (SEM) was adopted to analyze the data of the extended Technology Acceptance Model collected from the two groups of students who learned with the structured resource-based issue quest and the open resource-based issue quest approaches. The causal relationships of the modified model are shown in Figure 4, and the hypotheses are listed as follows.

- **Hypothesis H1:** The quality of the system causes perceived enjoyment.
- **Hypothesis H2:** The quality of the system results in perceived ease of use.
- **Hypothesis H3:** The perceived ease of use causes perceived usefulness.
- **Hypothesis H4:** Usefulness relates to enjoyment.
- **Hypothesis H5:** Perceived usefulness increases users’ intention to use the system.

![Figure 4. The SEM model of each group](image)

Measurement model

The measurement model was used to assess the reliability and validity of the observed variables, as shown in Table 4. The measurement models were evaluated based on three parts, the significance of each estimated coefficient or loading, the convergent validity, and the discriminant validity. Firstly, all items loaded significantly on their latent constructs \((p < .01)\). Hair et al. (2006) noted that an item is remarkable if its factor loading is greater than 0.50. In Table 4, the factor loadings (FL) of all the items in the measure range from 0.75 to 0.96, thus meeting the threshold (0.50), and demonstrating convergent validity at the item level.

Secondly, using composite reliability (CR) and average variance extracted (AVE) to evaluate the convergent validity, Fornell and Larcker (1981) suggested three phases for assessing convergent validity, including the item reliability of each measure, the composite reliability (CR) of each construct, and the average variance extracted (AVE). Nunnally and Bernstein (1994) noted that a value of .70 or higher is recommended to achieve adequate composite reliability (CR). The average variance extracted (AVE) is also an indicator of convergent validity. The
researchers have explained that the aim of the average variance extracted (AVE) is to measure the amount of variance captured by the construct in relation to the amount of variance attributable to the measurement error (Fornell & Larcker, 1981). A commonly used threshold value for average variance extracted (AVE) is 0.5 (Hair, Black, Babin, Anderson, & Tatham, 2006). As shown in Table 3, the convergent validity for the proposed constructs of this study is adequate when the value of AVE ranges from 0.66 to 0.88, exceeding 0.50. To sum up, the composite reliability (CR) and average variance extracted (AVE) both met the guidelines in Table 3. It is confirmed that the standardized coefficients show high validity and reliability in this study.

Table 4. Standardized coefficients for measuring the construct validity and reliability of the models of the two groups

<table>
<thead>
<tr>
<th>Latent variables (Construct)</th>
<th>Observed variables (Item)</th>
<th>Control group</th>
<th>Experimental group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor loading E</td>
<td>CR AVE</td>
<td>Factor loading ε</td>
</tr>
<tr>
<td>Quality of the Web</td>
<td>QOS1</td>
<td>0.89</td>
<td>0.21 0.95 0.87</td>
</tr>
<tr>
<td>Issue-Quest learning</td>
<td>QOS2</td>
<td>0.96</td>
<td>0.07</td>
</tr>
<tr>
<td>system (QOS)</td>
<td>QOS3</td>
<td>0.94</td>
<td>0.12</td>
</tr>
<tr>
<td>Perceived Enjoyment</td>
<td>PE1</td>
<td>0.90</td>
<td>0.18 0.93 0.82</td>
</tr>
<tr>
<td>(PE)</td>
<td>PE2</td>
<td>0.92</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>PE3</td>
<td>0.90</td>
<td>0.18</td>
</tr>
<tr>
<td>Users’ Intention (UI)</td>
<td>UI1</td>
<td>0.88</td>
<td>0.22 0.93 0.81</td>
</tr>
<tr>
<td></td>
<td>UI2</td>
<td>0.93</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>UI3</td>
<td>0.89</td>
<td>0.20</td>
</tr>
<tr>
<td>Perceived Usefulness (PU)</td>
<td>PU1</td>
<td>0.84</td>
<td>0.30 0.91 0.77</td>
</tr>
<tr>
<td></td>
<td>PU2</td>
<td>0.87</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>PU3</td>
<td>0.92</td>
<td>0.16</td>
</tr>
<tr>
<td>Perceived Ease of Use (PEU)</td>
<td>PEU1</td>
<td>0.87</td>
<td>0.24 0.86 0.66</td>
</tr>
<tr>
<td></td>
<td>PEU2</td>
<td>0.82</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>PEU3</td>
<td>0.75</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Thirdly, the correlational method needs to be used for determining the discriminant validity. Tables 5 and 6 show the implied correlations between the variables, which were all higher than 0.7. When the indicator is more correlated with the measured construct than with other constructs, discrimination can be achieved (Fornell, Tellis, & Zinkhan, 1982); moreover, two constructs could be highly correlated but remain distinct. In this study, discriminant validity was satisfactory for all operationalizations; that is, the constructs were adequate since discriminant validity was achieved at all the item and construct levels. Therefore, the scales could be considered to be satisfactory for employing the SEM analysis.

Table 5. Correlation between the variables in the control group (N = 204)

<table>
<thead>
<tr>
<th>Open resource-based issue quest</th>
<th>PE</th>
<th>UI</th>
<th>PU</th>
<th>PEU</th>
<th>QOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Enjoyment (PE)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Users’ Intention (UI)</td>
<td>0.74**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Usefulness (PU)</td>
<td>0.84**</td>
<td>0.88**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Ease of Use (PEU)</td>
<td>0.49**</td>
<td>0.54**</td>
<td>0.61**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Quality of the Web Issue-Quest learning system (QOS)</td>
<td>0.77**</td>
<td>0.63**</td>
<td>0.72**</td>
<td>0.64**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note. **p < .01.

Table 6. Correlation between the variables in the experimental group (N = 214)

<table>
<thead>
<tr>
<th>Structured resource-based issue quest</th>
<th>PE</th>
<th>UI</th>
<th>PU</th>
<th>PEU</th>
<th>QOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Enjoyment (PE)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Users’ Intention (UI)</td>
<td>0.77**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Usefulness (PU)</td>
<td>0.85**</td>
<td>0.91**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Ease of Use (PEU)</td>
<td>0.64**</td>
<td>0.70**</td>
<td>0.78**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Quality of the Web Issue-Quest learning system (QOS)</td>
<td>0.82**</td>
<td>0.72**</td>
<td>0.80**</td>
<td>0.77**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note. **p < .01.

Table 7 shows the descriptive statistics analyzing the distribution of the quantitative data obtained from the questionnaire, including the means and standard deviations of each scale for the control and experimental groups. The mean scores of QOS, PE, PEU, PU, and UI were all greater than 4.0, ranging from a low of 4.59 to a high of 4.96. This indicates an overall positive response to the constructs. Both the learners in the control group and in the experimental group had a positive attitude toward the PBL support system they used; therefore, the means of each scale for the two groups are high and similar, so there are no remarkable differences between the
descriptive statistics of the two groups. The students in the experimental group have a similarly strong degree of perception as the students in the control group.

Table 7. The investigation results of each scale for the two groups

<table>
<thead>
<tr>
<th>Scale</th>
<th>Control group (N = 204)</th>
<th>Experimental group (N = 214)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Quality of the Web Issue-Quest learning system (QOS)</td>
<td>4.92</td>
<td>1.19</td>
</tr>
<tr>
<td>Perceived Enjoyment (PE)</td>
<td>4.59</td>
<td>1.27</td>
</tr>
<tr>
<td>Perceived Ease of Use (PEU)</td>
<td>4.89</td>
<td>1.00</td>
</tr>
<tr>
<td>Perceived Usefulness (PU)</td>
<td>4.96</td>
<td>1.16</td>
</tr>
<tr>
<td>Users’ intention (UI)</td>
<td>4.93</td>
<td>1.13</td>
</tr>
</tbody>
</table>

Structural model

The fit measures of the two-group structural model and correlation of the scales in both groups are shown in Table 8. For most fit indices, the fit values were remarkably good. Only the fit index values of RMSEA and Standardized RMR were in the “acceptable” range. Consequently, an acceptable model fit was derived.

It was also found that all of the correlation coefficients were positive and less than 1.0, implying that all of the hypotheses related to the construct-to-construct relationships were accepted.

Table 8. Fit measures for the two-group structural model

<table>
<thead>
<tr>
<th>Fit index</th>
<th>Fit values for two-groups</th>
<th>Recommended value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degrees of Freedom (DF)</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>Minimum Fit Function Chi-Square ($\chi^2$)</td>
<td>558.02 ($p = .00$)</td>
<td></td>
</tr>
<tr>
<td>control group = 258.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>experimental group = 299.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$/DF</td>
<td>3.39</td>
<td>&lt; 5.0</td>
</tr>
<tr>
<td>Percentage Contribution to Chi-Square</td>
<td>Control group = 46.26%</td>
<td></td>
</tr>
<tr>
<td>Experimental group = 53.74%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root Mean Square Error of Approximation (RMSEA)</td>
<td>0.095</td>
<td>&lt; 0.08</td>
</tr>
<tr>
<td>90 Percent Confidence Interval for RMSEA</td>
<td>(0.086 ; 0.11)</td>
<td></td>
</tr>
<tr>
<td>P-Value for Test of Close Fit (RMSEA &lt; 0.05)</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Goodness of Fit Index (GFI)</td>
<td>0.86</td>
<td>≥ 0.90</td>
</tr>
<tr>
<td>Standardized RMR</td>
<td>0.084</td>
<td>≤ 0.05</td>
</tr>
<tr>
<td>Normed Fit Index (NFI)</td>
<td>0.97</td>
<td>&gt; 0.9</td>
</tr>
<tr>
<td>Non-Normed Fit Index (NNFI)</td>
<td>0.98</td>
<td>&gt; 0.9</td>
</tr>
<tr>
<td>Comparative Fit Index (CFI)</td>
<td>0.98</td>
<td>&gt; 0.9</td>
</tr>
<tr>
<td>Relative Fit Index (RFI)</td>
<td>0.97</td>
<td>&gt; 0.9</td>
</tr>
<tr>
<td>Incremental Fit Index (IFI)</td>
<td>0.98</td>
<td>&gt; 0.9</td>
</tr>
<tr>
<td>Parsimony Normed Fit Index (PNFI)</td>
<td>0.81</td>
<td>&gt; 0.8</td>
</tr>
</tbody>
</table>

To sum up, this study tested the hypothesized model modified from the model proposed by Liaw and Huang (2003) and from the literature; therefore, the findings have theoretical underpinnings. Based on the analysis results in Table 8, the hypotheses of the direct relationships H1, H2, H3, H4, and H5 in Figure 4 were all supported. Accordingly, Figure 5 and Figure 6 reveal the obtained results of the hypothesized structural mode for the experimental group and the control group.
The following concludes the two structure models. The external variable, the quality of the Web Issue-Quest learning system (QOS), in the structured resource-based issue quest mode (i.e., the experimental group) had strong causal relations with perceived enjoyment (PE) and perceived ease of use (PEU). In the two constructs, it was found that perceived enjoyment (PE) had similar causal relations with perceived usefulness (PU) in the two groups. In particular, perceived ease of use (PEU) in the structured resource-based issue quest mode (i.e., the experimental group) in Figure 6 had a stronger causal relation to perceived usefulness (PU) than the open resource-based mode; on the other hand, the open resource-based issue quest mode showed a relatively stronger causal relation from perceived usefulness (PU) to intention to use (IU).

**Discussion and conclusions**

In this study, a structured Web issue-quest approach was proposed and applied to a high school computer programming course. The experimental results showed that the students who learned with the proposed approach had significantly better learning achievements than those who learned with the conventional approach. This finding conforms to the report of Riahinia and Zandian (2008) who indicated that a large proportion of students (63.4%) preferred accessing well-structured learning resources (e.g., online databases or e-libraries) than searching for information on the Internet when they were novices in a field. Therefore, it is suggested that, to improve students’ learning efficiency for specified issues, teachers can consider adopting the structured resource-based issue quest (i.e., the information searching system in the experimental group).

Meanwhile, it was found that the students’ perceived enjoyment, perceived ease of use, and perceived usefulness were key determinants of behavioral intention. These results agree with the finding of Giannakos and Vlamos (2013) who investigated the acceptance of educational websites and suggested that instructors and educational institutions should focus on usefulness and ease of use. The results of this current study also showed that the students had strong perceptions of the ease of use and the usefulness of the Web Issue-Quest activity, no matter whether they used the open resource-based or the structured resource-based mode for learning the basic concepts of structured programming. That is, both modes of Web Issue-Quest learning engaged the students in learning with high intention.

On the other hand, it was found that the structured resource-based Web Issue-Quest method was more effective than the open resource-based method for the students’ learning achievement, but no difference was found between the two methods in terms of their perceptions. One of the reasons for obtaining such opposite results could be due to the fact that the interfaces for both systems were identical, although the sources of the searched data were totally different. Such a finding is similar to what was reported by Hwang, Sung, Hung, and Huang (2013), who found that most students’ preferences regarding using computer systems were highly related to the system interfaces.

To sum up, the main contribution of this study is proposing and showing the effectiveness of engaging students in structured resource-based Web Issue-Quest environments for computer programming courses. From the analysis results of the students’ learning achievements and perceptions, it was found that the structured resource-based method was more beneficial to the students than the open resource-based method; meanwhile, the students had equivalent perceptions of the two methods. These findings provide a good reference for those teachers and researchers who intend to conduct Web Issue-Quest learning activities based on solid evidence obtained by conscientious data analysis. That is, the provision of a structured data source is important for improving students’ learning outcomes, and the design of the user interface could be the major factor affecting their perceptions of using the learning system.

On the other hand, the research limitation of this study needs to be noted. In this study, the learning approach was designed by taking the needs of novice learners into account, and hence the findings might not be able to be
inferred to all learners. In the future, it would be worth conducting additional studies to further investigate the impacts of the proposed approach on students with different personal factors, such as computer experience, knowledge levels and learning styles. It could also be interesting to take different user interface strategies into consideration.

Acknowledgements

This study is supported in part by the National Science Council of Taiwan under contract numbers NSC 105-2628-S-003-002-MY3, MOST 103-2628-S-003-003-MY2, and NSC 102-2511-S-011-007-MY3. This work is also partly supported by “Aim for the Top University Project” of the National Taiwan Normal University and the Ministry of Education, Taiwan, R.O.C.

References


Student Engagement in Long-Term Collaborative EFL Storytelling Activities: An Analysis of Learners with English Proficiency Differences

Yun-Yin Huang1, Chen-Chung Liu2*, Yu Wang2, Chin-Chung Tsai3 and Hung-Ming Lin4

1National Tsing Hua University, Taiwan // 2Graduate Institute of Network Learning Technology, National Central University, Taiwan // 3Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // 4Department of Business Administration, Minghsin University of Science and Technology, Taiwan // yuyhuang@cl.ncu.edu.tw // ccliu@cl.ncu.edu.tw // fish7588@gmail.com // cctsa@mail.ntust.edu.tw // hmlin@must.edu.tw

*Corresponding author

(Submitted January 19, 2016; Revised June 1, 2016; Accepted September 8, 2016)

ABSTRACT

English proficiency difference among students is a challenging pedagogical issue in EFL classrooms worldwide. Collaborative digital storytelling has been adopted in language learning settings to increase motivation and engagement, especially for young learners. However, it remains unknown whether students of different proficiency levels can equally benefit from this collaborative approach. Thus, this study implemented a 17-month technology-enhanced collaborative storytelling activity and examined young students’ pair performance, flow perception, and learning strategies in relation to students’ English proficiency level. The students’ proficiency level was found to be an influencing factor of their engagement patterns, use of learning strategies, and pair performance. These findings support the low-threshold-high-ceiling principle, suggesting that collaborative activities should ensure students of different proficiency share the same goals, while allowing different types of participation to maximize their engagement. It is hoped that the findings and pedagogical suggestions can address the issue of proficiency differences in EFL classrooms and serve as a reference for future research of EFL collaborative storytelling activities.

Keywords

EFL, Storytelling, Engagement, Collaboration, Proficiency

Introduction

English proficiency is considered a critical competence in today’s increasingly globalized society. In many Asian countries, English language education was introduced into primary schools years ago (Chen, 2013). However, owing to students’ various access levels to resources, English proficiency differences among students are becoming apparent (Baldauf, Kaplan, Kamwangamalu, & Bryant, 2011), meaning that students come to school with significant differences in their English proficiency. Such a difference has become a critical pedagogical concern (Chang, 2006). English as Foreign Language (EFL) teachers, particularly in elementary education, are facing significant pedagogical challenges to cater to individual students’ needs (Nunan, 2003) such as the adoption of various language learning strategies (Uhl Chamot & El-Dinary, 1999) and different proficiency levels (Wang, 2008). It is often observed that when the teacher targets and designs his/her instruction for a certain proficiency level, the more or less proficient students in the class would be neglected and thus become disengaged. In particular, language learning is a long-term process in which learners’ motivation and engagement change over time, and engaging learners of various proficiency levels in language activities of extended periods of time is a critical challenge and research focus.

Engagement, as a multi-dimensional factor, has been reported to be associated with students’ enjoyment, motivation, confidence, perceived usefulness, performance, and flow perceptions in technology-enhanced language learning (Liu, Wang, & Tai, 2016). Various technologies have been adopted in language learning settings to facilitate the learning of students of different proficiency levels, and to increase engagement to overcome this challenge. Digital storytelling has been reported to effectively foster young EFL learners’ interest in learning (Figg & McCartney, 2010). However, it remains unknown whether students of different proficiency levels can equally benefit from digital storytelling. Students’ dynamic and changing patterns of engagement during the learning process are also unknown. Thus, in this study, we implemented a long-term technology-enhanced collaborative storytelling activity and examined young students’ long-term flow perception, use of learning strategies, and pair performance in relation to their English proficiency level. Data were collected from 30 collaborative storytelling sessions over the 17-month period, so as to identify the changing patterns in student engagement (Herrington, Oliver, & Reeves, 2003; Liu et al., 2016). Student engagement was examined according to the flow theoretical framework (Csikszentmihalyi, 1975) and the Strategy Inventory for Language Learning (Lan, 2005; Oxford, 1990) to better understand the interplay of the students’ engagement, use of learning strategies, and pair performance in the learning activity.
Literature review

Digital storytelling

Storytelling is an effective teaching strategy for young EFL learners, and the integration of technologies in storytelling instruction provides students with opportunities to synthesize verbal and imagery representations based on their initial understanding (Lotherington & Jenson, 2011). Digital storytelling (DST) “takes the ancient art of oral storytelling and engages a palette of technical tools to weave personal tales using images, graphics, music, and sound mixed together with the author’s own story voice” (Porter, 2005). The literature has confirmed that DST can improve multiple language skills, including sentence construction ability (Kim, 2014; Tsou, Wang & Tzeng, 2006), writing (Figg & McCartney, 2010; Yang & Wu, 2012; Yoon, 2013), listening comprehension (Yoon, 2013), and oral skills (Tahriri, Tous, & Movahedfar, 2015). In addition to the benefits which traditional storytelling could bring, the multimedia feature of digital storytelling has been reported to be beneficial in terms of capturing attention, fostering interest to learn, and enhancing engagement in learning activities (Di Blas, Garzotto, Paolini, & Sabiescu, 2009; Figg & McCartney, 2010; Sadik, 2008). With the Internet connection, digital storytelling allows students to not only extend their imagination through digital tools, but also to share their stories with a wider audience (Kim, 2014; Liu, Lin, Deng, & Tsai, 2014).

The collaborative approach has been considered an effective alternative for teaching students of various proficiency levels (Nunan, 1998), and has been combined with DST to enable students to work together to develop and organize multimedia materials for storytelling (e.g., Gelmini-Hornsby, Ainsworth, & O’Malley, 2011; Liu, Tao, Chen, Liu, & Chen, 2013). Collaborative DST can help to facilitate a reciprocal learning process in which students play different roles and learn from each other to construct a story together (Liu, Liu, Wang, Chen, & Su, 2012). When synthesizing ideas and creating a shared story, students can foster critical thinking through communication and enhance their creativity (Nordmark & Milrad, 2012; Yang & Wu, 2012). However, with all of these educational potentials that collaborative DST could bring, it is noted that working together does not automatically engage students in productive construction (Gelmini-Hornsby et al., 2011; Kreijns, Kirschner, & Jochems, 2003; Liu & Tsai, 2008). Past research has confirmed positive perceptions of enjoyment and satisfaction when technologies were used to support collaboration (Asoodar et al., 2014; Ducate et al., 2011). However, few studies have attempted to probe engagement from theoretical perspectives or to examine its changing patterns from longitudinal observations. Students with limited language knowledge have been found to have difficulties producing quality language work in collaborative DST activities, and often experience several phases of disengagement over an extended period of time (Liu, Wang, & Tai, 2016). It is thus worthwhile to investigate long-term engagement patterns in collaborative digital storytelling activities.

Long-term engagement

Engagement can be viewed as active participation in the learning process, and contributes to deeper and more meaningful learning. When the learners are involved and interested in meaningful tasks, they learn more effectively, and are more likely to retain the information and transfer it to other contexts (Kearsley & Schneiderman, 1998). Engagement is a multifaceted construct characterized by various dimensions such as challenge, sensory appeal, attention, feedback, curiosity, and interest (O’Brien & Toms, 2008; Trevino & Webster, 1992). Engagement has frequently been associated with flow perceptions (Csikszentmihalyi, 1975), both of which share overlapping constructs. Flow is the state in which people are deeply involved in an activity and enjoy the experience for its own sake rather than for any other reason; it shares a few attributes with engagement, including focused attention, feedback, control, and intrinsic motivation (Csikszentmihalyi, 1975). Flow perceptions have been found to be positively associated with exploratory and participatory activities, and students often demonstrate multiple strategies when in a flow state (Liu, Cheng, & Huang, 2011; Liu, Wang, & Tai, 2016).

Ideally, it is hoped that students could continuously engage in and apply more learning strategies in a flow state for an extended period of time for better and lasting language learning performance. However, discrepancies in the prior language proficiency of students have been found to affect engagement in technology-enhanced activities, and to lead to the use of different language learning strategies. High proficiency learners are more likely to effectively use a combination of multiple strategies, and lower proficiency learners tend to apply one strategy at a time (Uhl Chamot & El-Dinary, 1999; Oxford, 1990). While proficiency level is an influencing factor of student engagement and language learning strategy use in conventional instruction, it is unknown if this is true in technology-enhanced collaborative settings in which dealing with proficiency differences might become an even more challenging pedagogical task.
To fill this gap in the research literature and teaching practice, this study investigated how student pairs of different language proficiency levels engaged in EFL collaborative digital storytelling activities, and examined their engagement patterns, language learning strategies, and storytelling performance specifically. This study attempted to answer the following questions:

- **RQ 1:** Did students’ flow perceptions of the collaborative storytelling activity present specific patterns? If so, was students’ prior English proficiency an influencing factor of their flow patterns?
- **RQ 2:** Did students of different English proficiency use learning strategies differently in the collaborative storytelling activity?
- **RQ 3:** How did student pairs of different English proficiency levels perform in the collaborative storytelling activity?

**Method**

This study investigated the interrelationships between students’ entry English proficiency, flow perception, and language learning strategies in the context of the collaborative DST (Figure 1). Owing to the complex nature and the variety of variables involved with engagement, a mixed-methods approach (Creswell & Clark, 2011) was adopted for a more comprehensive and holistic understanding of the complex learning context. The mixed methods approach integrates quantitative and qualitative data to supplement the interpretation of different data types (or the limitations of one type of data are balanced by the strengths of another). Analysis of both quantitative and qualitative data has been adopted in similar children’s storytelling settings (e.g., Yoon, 2013). Following the sequential explanatory design, follow-up student interviews were conducted after the statistical analysis of survey data in order to explain the initial findings and to further explore the nuances of the learning process.

**Participants**

The participating students of this study were 42 third graders from two classes in an urban elementary school in northern Taiwan. Their ages ranged from 9 to 10 years old. These students were still in the early stages of English language development, and their learning focused on basic vocabulary and sentence patterns. The 42 students were grouped into 21 pairs for a collaborative digital storytelling activity based on their entry English proficiency. Proficiency assessment included the school’s standard midterm examination and oral tests for reading fluency and vocabulary knowledge. The materials used in oral reading fluency and vocabulary tests were adopted from the target content (15 stories from Starfall®) of the collaborative storytelling activity. The vocabulary part was a 20-item English - Chinese meaning match test. The oral reading fluency part was 15 sentences (76 words) retrieved from the same material. The number of words that each student could read aloud correctly per minute was counted as an indicator of their oral reading ability.

This study computed the grades of the midterm examination, oral reading fluency, and vocabulary knowledge to represent the students’ holistic English proficiency. The students were divided into three different student groups (high-mid-low proficiency) based on the overall computed scores. The majority (22) of the students were rated with a score between ± .5 standard deviation, as mid-proficient students, due to the normal distribution of students’ proficiency level. Another 8 students scored .5 standard deviation over the mean and so were considered as high-proficient students, while the remaining 12 students who were rated with a grade .5 standard deviation were considered as low-proficient students.
deviation below the mean were considered the low-proficient students. Students were divided into pairs on a random basis with one exception being that high-proficient students were not grouped together.

**Procedures**

The implementation of the collaborative digital storytelling activity was 17 months (1.5 hours a week). The activity was approved and supported by the homeroom teacher and the school authority. We originally aimed to implement the activity every week, but due to school holidays and school-wide examinations, a few weeks were skipped during the 17-month period. Participating students finally completed a total of 30 sessions in 30 weeks. Proficiency assessment was conducted before the activity implementation. In each activity session, each pair worked to retell a story with drawings and oral narration, and published their story on a Web 2.0 platform. After each activity session, the students answered a flow survey to understand their flow perceptions. Therefore, a sequence of 30 sets of flow survey results was obtained. Upon completion of the 30 sessions of the activity, a strategy inventory for language learning was administered to elicit the students’ language learning strategies.

**The collaborative digital storytelling activity**

The goal of the digital storytelling activity was to create a multimedia picture book by re-telling model stories, using drawings, texts, and audio-recording. Throughout the 17-month period of this study, the pairs of participants were guided to use iPads to collaboratively re-tell stories using their own English sentences and drawing on a weekly basis. In other words, the participating students sat physically together to share one iPad, and they needed to read and practice reading aloud the model stories, draw the scenes in the story (Figure 2), and record their oral reading to create a picture book. We did not implement a strict timeframe but allowed the students to re-tell stories and create picture books at their own pace. Some pairs therefore completed more stories than others.

![Figure 2. Students using the storytelling app to create a multimedia story](image)

Collaborative digital storytelling for young children is a challenging task, so in this study we implemented the activity over the 17 months and allowed the students sufficient time to become familiar with the technology. Moreover, we provided model stories as scaffolding in the early stage of the intervention to help the students develop the skills and knowledge required for the task. An e-book of 15 model stories (Starfall®), including
graphs and texts along with oral narrations, was installed on the iPads as scaffolding in the early stage of the story re-telling. Upon completing a story, the students shared it on an online platform *Story & Painting House* (Figure 3), on which they could publish their self-created multimedia picture books and also view others’ work. The students saved their progress every week on the platform, and continued to work on their stories the next time. When completing a story, the student pairs had opportunities to showcase their picture books on stage in front of the class.

**Data collection**

Based on the research framework (Figure 1), we implemented surveys and interviews to investigate the students’ flow perceptions and language learning strategies. Furthermore, we examined the student artifacts (multimedia picture books) to assess each student pair’s performance in the activity. These data sources were triangulated with follow-up interviews at multiple times to achieve a more comprehensive understanding of the students’ overall engagement in the activity.

**Flow survey**

To measure individuals’ flow perceptions, we adopted Trevino and Webster’s (1992) flow model for our survey, asking about the four dimensions of flow perceptions: control, attention, curiosity, and intrinsic interest in four items on a 5-point Likert scale (5 being the highest; 1 being the lowest flow perception). The perceptions of these four dimensions together were combined and analyzed as general flow perceptions. Owing to the young age of the participating students, every survey question was read aloud by the homeroom teacher to ensure their understanding, and then the students checked the fitting statement with pencils on a printed form. In between each question, the teacher would pause and confirm all students’ understanding and completion before moving on to the next question. The flow survey was administered immediately after each storytelling activity, making it a total of 30 times. We used the average scores of the students’ responses to present their overall flow perception. The Cronbach’s reliability (alpha) of the four items used in this study was .75, showing adequate reliability of the survey items.

**Strategy Inventory for Language Learning (SILL)**

To understand the students’ use of learning strategies, we adopted the *Children’s Strategy Inventory for Language Learning* (SILL) (Lan, 2005), revised from the original 50-item SILL (Oxford, 1990); both measurements have good established validity and have been widely used to assess language learning strategies. The Children’s SILL used in this study consists of 30 items investigating students’ application of six learning strategies: cognitive, memory, metacognitive, compensatory, affective, and social strategies. For example, items for cognitive strategies include “(w)hen I speak in English, I try to imitate English-speaking people, in order to pronounce the words correctly,” and “I make an effort to understand the sense of what I read or what I hear without translating word for word.” Items for metacognitive strategies include: “I organize my time to study English,” and “I analyze the errors I have made and try not to repeat them.” The items use a 5-point Likert scale, 1 indicating the lowest level of strategy application and 5 indicating the highest level. The administration of SILL was identical to that of the Flow survey, and was conducted at the end of the 17-month period.

**Multimedia picture books and pair performance in the collaborative storytelling activity**

In the digital storytelling activity, students produced multimedia picture books by recording their oral reading and drawing on iPads. The picture books produced by the students were analyzed according to the following three components:

- **Language productivity** was the amount of target language output, including (1) the number of stories each pair created, (2) the average number of pages in each story, (3) the number of English sentences in each scene, and (4) the number of audio-recorded lines per story. Some student work contained more pages for each story, and more detailed drawing that matched and enhanced the storyline, while some other works included only sketches of a few lines. These numerical data were directly retrieved from the *Story & Painting House* system log.
- **Drawing presentation** was the pictorial product of the students’ multimedia picture books, which was meant to match and enhance the storyline. Two research assistants analyzed and rated the drawing presentation; they provided not only numerical scores but also substantial descriptions of the pictorial and linguistic features of the students’ work in the hope of identifying patterns in their drawing presentations. While analyzing the drawings, the two raters listed the major features of the multimedia picture books and assigned scores, from 1 as the lowest to 5 as the highest, based on the extent to which the drawing matched and enhanced the storyline. An acceptable inter-rater agreement was reached (r = .75, p < .01) between the two raters’ scores.

- **Audio recording** is another component of students’ multimedia picture books. The students were not instructed specifically on how to record their oral reading; their recording data could therefore be one way to show their pair collaboration pattern.

**Follow-up interviews**

In order to gauge students’ views and experience in the collaborative storytelling activity, we conducted 26 interviews with a focus on their flow perceptions (RQ1) and use of language learning strategies (RQ2) to better understand their changing flow perceptions and various use of language learning strategies. Example initial interview questions included, “Did you feel bored during the activity, and why?” “How did you figure out and learn the words you didn’t know in the model story?” and “How did you and your partner collaborate to create a story?” These initial questions were designed to be as open as possible. It was hoped that the unstructured nature of the conversation would help to probe the students’ personal experiences during the activity.

**Data analysis**

To answer the first research question, ANOVA with repeated measures was applied to analyze students’ flow perceptions over the 30 activity sessions. To obtain a comprehensive result, we calculated the average flow perceptions of the first 10 times, the middle 10 times, and the last 10 times to represent the students’ flow perceptions during the initial, middle, and ending phases of the activity. We then applied ANOVA with repeated measures to compare the flow patterns during the three activity phases with students’ English proficiency (high – mid – low) serving as the between-subjects factor. ANOVA with repeated measures can identify whether there are significant differences in the patterns of flow perceptions among the three student groups.

To answer the second research question, we conducted one-way ANOVA with students’ English proficiency (high – mid – low) serving as independent variables for examining the differences in the applications of the six categories of language learning strategies. Furthermore, to answer the third research question regarding students’ pair performance, we firstly examined the drawing presentation and the English sentences in the student-created multimedia picture books (artifact), and then conducted a K-means cluster analysis to identify major student pairs which demonstrated similar pair performance based on the two aforementioned aspects. The pair combinations of each cluster were then further interpreted together with the initial descriptive data of the students’ work and two raters’ observations of their audio recording patterns.

**Results**

The results include the findings from the analysis of the flow patterns (RQ1), use of learning strategies (RQ2), and pair performance (RQ3) in the collaborative storytelling activity with a focus on the influences of difference proficiency (low-, mid- and high-proficiency) in alignment with the research framework (Figure 1).

**Flow patterns**

In this study we conducted an ANOVA with students’ English proficiency (high – mid – low) serving as independent variables to examine students’ flow perception during the initial, middle and ending phases of the collaborative storytelling activity. It should be noted that three of the 42 students did not complete all of the required tests and questionnaires. Therefore, the results involved only the data from the remaining 39 students.
Table 1. ANOVA analysis of the flow perceptions of the three student groups

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>Phase</th>
<th>M</th>
<th>SD</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>High (N = 8)</td>
<td>Initial</td>
<td>4.49</td>
<td>.54</td>
<td>4.28**</td>
<td>&lt;.01</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>4.23</td>
<td>.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ending</td>
<td>4.03</td>
<td>.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid (N = 22)</td>
<td>Initial</td>
<td>4.06</td>
<td>.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>4.30</td>
<td>.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ending</td>
<td>4.42</td>
<td>.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (N = 9)</td>
<td>Initial</td>
<td>4.17</td>
<td>.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>4.18</td>
<td>.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ending</td>
<td>4.43</td>
<td>.57</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. **p < .01.

The low-, mid- and high-proficiency students demonstrated significantly different flow patterns (F = 4.28, p < .01), showing different change trends in flow perceptions during the three phases of the activity (Table 1). The flow perceptions of the mid- and low-proficiency students rose gradually, while those of the high-proficiency group dropped during the period of this study (Figure 4). Taking a closer look, we can see that both the mid- and low-proficient student groups perceived lower flow perceptions in the initial phase than the high-proficiency group. However, their flow perceptions increased significantly during the middle and ending phases. More specifically, the flow perceptions of the low-proficient students slightly increased during the middle phase, and then skyrocketed in the ending phase. On the other hand, the flow perceptions of the mid-proficient students increased dramatically first, and then slowed down but continued to increase gradually in the second half of the study. The most noticeable flow trend we observed is that the flow perceptions of the high-proficiency group decreased steadily throughout the activity.

Figure 4. The flow perception patterns of the three student groups (high-mid-low proficiency)

The ANOVA analysis with repeated measures was also applied to each dimension of the students’ flow perceptions. The results shown in Table 2 indicate that the three student groups showed significantly different patterns in all four dimensions of flow perceptions: attention (F = 3.31, p < .05), curiosity (F = 3.38, p < .05), control (F = 4.42, p < .01), and interest (F = 2.89, p < .05). High-proficiency students’ perceptions of attention, curiosity, control, and interest decreased during the three activity phases, while the mid-proficiency students’ flow perceptions in all four dimensions gradually increased during these activity phases. However, the low-proficiency students reported another different trend in their control and interest perceptions, which involved a decrease during the middle phase and an increase in the ending phase of the activity. The flow perceptions of the low-proficiency students might have resulted from their lack of the English skills needed to effectively participate in the activity and the lack of interest in the activity until the middle phase. It is noted that the low-proficiency students soon gained higher-level perceptions of control and interest after the middle phase of the activity.
his study—e 5). Nine student groups exhibited the symbolic and imitative styles. The works of both symbolic and imitative styles were still mainly based on the model. student works of both symbolic and imitative styles were still mainly based on the model stories and demonstrated more of the symbolic style in those in the model stories. It was found that presentation, we l...

Artifact analysis and pair storytelling performance

To understand the students’ digital storytelling performance in pairs, we analyzed the picture books in terms of their (1) language productivity, (2) drawing presentations, and (3) audio recordings. As described earlier, data of student language productivity was retrieved from the system log, and the analysis of the students’ drawings and audio recordings was conducted by two independent coders.

We analyzed a total of 146 picture books produced by the students during the study period. In terms of drawing presentation, we looked into the expression of the characters, the brushwork, and the relations of the drawing to those in the model stories. It was found that the students demonstrated three distinct styles of drawing presentation: symbolic style (Figure 5), imitative style (Figure 6), and transformative style (Figure 7). Eight pairs demonstrated more of the symbolic style in their self-created picture books, as they used only simplified lines to represent the scenes of the model stories and to retell the model story (Figure 5). Nine student groups exhibited an imitative style, consisting of well-defined characters, detailed and colorful backgrounds, and delicate brushwork (Figure 6). Student works of both symbolic and imitative styles were still mainly based on the model stories in terms of composition, and the setting and design of the characters. The pairs who adopted a transformative style tended to add a twist to their drawing presentation to accompany their stories. For example, a female pair changed the original character in the model story from a boy to a girl in their drawing so as to better present themselves (Figure 7). Based on the extent of how the drawing presentations matched and enhanced the storyline, the student works which adopted transformative and imitative styles were given higher scores (4–5), and those using the symbolic style were marked lower (2–3). It has been noted that students’

---

**Table 2. Four flow dimensions of the three proficiency groups**

<table>
<thead>
<tr>
<th>Proficiency level</th>
<th>Activity phases</th>
<th>Attention</th>
<th>Curiosity</th>
<th>Control</th>
<th>Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>High (N = 8)</td>
<td>Initial</td>
<td>4.50</td>
<td>0.48</td>
<td>3.31**</td>
<td>4.55</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>4.20</td>
<td>0.81</td>
<td>3.84**</td>
<td>4.43</td>
</tr>
<tr>
<td></td>
<td>Ending</td>
<td>4.13</td>
<td>0.75</td>
<td>3.78**</td>
<td>4.10</td>
</tr>
<tr>
<td>Mid (N = 22)</td>
<td>Initial</td>
<td>3.93</td>
<td>0.74</td>
<td>2.92</td>
<td>4.16</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>4.18</td>
<td>0.68</td>
<td>3.83**</td>
<td>4.34</td>
</tr>
<tr>
<td></td>
<td>Ending</td>
<td>4.29</td>
<td>0.64</td>
<td>3.69</td>
<td>4.47</td>
</tr>
<tr>
<td>Low (N = 9)</td>
<td>Initial</td>
<td>4.11</td>
<td>0.73</td>
<td>2.75**</td>
<td>4.08</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>4.13</td>
<td>0.92</td>
<td>2.67</td>
<td>4.18</td>
</tr>
<tr>
<td></td>
<td>Ending</td>
<td>4.36</td>
<td>0.73</td>
<td>3.82**</td>
<td>4.38</td>
</tr>
</tbody>
</table>

*Note. *p = .05; **p = .01.

**Table 3. The difference between the high-, mid-, and low-proficient students in their use of language learning strategies**

<table>
<thead>
<tr>
<th>Strategy category</th>
<th>High (N = 8)</th>
<th>Mid (N = 22)</th>
<th>Low (N = 9)</th>
<th>F</th>
<th>p</th>
<th>Post-hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>2.69 (.85)</td>
<td>3.80 (.91)</td>
<td>3.33 (1.04)</td>
<td>4.29*</td>
<td>.02</td>
<td>Mid &gt; High</td>
</tr>
<tr>
<td>Cognitive</td>
<td>3.00 (.82)</td>
<td>3.29 (.99)</td>
<td>2.80 (.94)</td>
<td>0.94</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Compensation</td>
<td>2.75 (.77)</td>
<td>3.51 (1.00)</td>
<td>2.83 (.65)</td>
<td>3.16</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Metacognitive</td>
<td>2.93 (1.12)</td>
<td>3.55 (.98)</td>
<td>2.71 (.68)</td>
<td>2.95</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Affective</td>
<td>2.71 (.99)</td>
<td>3.44 (1.03)</td>
<td>2.96 (.75)</td>
<td>1.96</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>2.63 (.86)</td>
<td>3.38 (.92)</td>
<td>2.67 (1.05)</td>
<td>2.91</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>2.83 (.70)</td>
<td>3.48 (.85)</td>
<td>2.90 (.79)</td>
<td>2.84</td>
<td>.07</td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p < .05.
drawing presentation and language productivity are two different dimensions of their multimedia creation that support and complement each other. Thus, we examined not only the linguistic but also the pictorial aspects of the artifacts so as to present a more holistic view of their pair performance in their collaborative digital storytelling.

After the initial analysis of student artifacts, student pair performance in the collaborative storytelling activity was analyzed using the K-means cluster technique based on their language productivity and drawing presentation. The results identified three main clusters, while two pairs were identified as outliers as they could not be classified into any of the three main clusters. Clusters were named based on their general performance in their language productivity and drawing presentation performance as high language productivity pairs (HLPP), creativity-oriented pairs (COP), and low performing pairs (LPP) (Table 4). The result of the MANOVA shows that there was a statistically significant difference in pair performance between the three clusters ($F = 13.78, p < .01$; Wilk's lambda=.02; partial $\eta^2 = .852$) indicating that the between-group means were all reliably distinguished. The activity performances of the three clusters and their relation to the groups' English proficiency are detailed below:

<table>
<thead>
<tr>
<th>Clusters (No. of pairs)</th>
<th>Stories</th>
<th>Pages</th>
<th>Sentences</th>
<th>Recordings</th>
<th>Drawing scores</th>
<th>Pair combination (No. of pairs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLPP (5)</td>
<td>9.6</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
<td>3.25</td>
<td>MH(3), LM(1), LH(1)</td>
</tr>
<tr>
<td>COP (9)</td>
<td>7.33</td>
<td>7.26</td>
<td>7.4</td>
<td>7.39</td>
<td>3.44</td>
<td>MM(3), LM(3), MH(2), LH(1)</td>
</tr>
<tr>
<td>LPP (5)</td>
<td>6.4</td>
<td>5.95</td>
<td>5.57</td>
<td>5.47</td>
<td>2.76</td>
<td>LM(2), MM(2), LL(1)</td>
</tr>
<tr>
<td>Outlier 1</td>
<td>8</td>
<td>5.38</td>
<td>5.25</td>
<td>5.25</td>
<td>1.94</td>
<td>HL(1)</td>
</tr>
<tr>
<td>Outlier 2</td>
<td>10</td>
<td>4.6</td>
<td>4.5</td>
<td>4.5</td>
<td>2.55</td>
<td>ML(1)</td>
</tr>
</tbody>
</table>

*Pair combination: L (low-proficient students), M (mid-proficient students), H (high-proficient students).
High language productivity pairs (HLPP, N = 5): The first cluster was the “high language productivity pairs” because the pairs in this cluster generally had the most significant language productivity (numbers of stories, pages, English sentences, and audio-recordings). In other words, the pairs in this cluster produced more stories than the other pairs, and the number of sentences and pages in their stories were also significantly more than those created by the other pairs. On average, these high performing pairs produced 9.6 stories during the period of this activity, and each story contained 7.5 pages and 7.5 sentences. Their average number of recorded sentences was 7.5 and their drawing score was 3.25.

Creativity-oriented pairs (COP, N = 9): The second cluster was the “creativity-oriented pairs,” identified as the majority in our study. Nine student pairs belonged to this cluster. The student pairs in this cluster obtained the highest scores in their creativity-oriented performance, that is, the drawing scores, but showed less language productivity than the HLPP pairs. Their average drawing score was 3.44, which was higher than that of the other two clusters. However, their language productivity was only at the middle level. On average, these middle performing pairs produced 7.33 stories, each of which consisted of 7.26 pages and 7.4 sentences. Their average number of recorded sentences was 7.39. The pair combinations of this cluster were varied. Some of the pair combinations were students of different English proficiencies; some were of the same proficiency.

Low performing pairs (LPP, N = 5): The third cluster was “low performing pairs” because the student pairs in this cluster demonstrated neither a high level of language productivity nor creativity-based performance. They produced significantly fewer sentences and stories than the high performing pairs. On average, these low performing pairs produced 6.4 stories in the activity, each consisting of 5.95 pages and 5.57 sentences. Their average number of recorded sentences was 5.47, and their drawing score was 2.76. It is noted that there were no high-proficient students in any of these pairs.

We also conducted an ANOVA analysis of the flow perceptions of the HLPP, COP, and LPP pairs, and the results indicated no significant difference between any of these three pair groups. In other words, students’ flow perceptions were not directly influenced by their pair storytelling performance. The students who produced less satisfactory performance (LPP) did not perceive significantly higher or lower flow perceptions than the HLPP or COP pairs (F = 0.22, p = .90). However, there was a noteworthy relationship between English proficiency and pair storytelling performance, with the three clusters exhibiting the following characteristics:

- Most of the pairs including a high-proficient student demonstrated satisfactory performance. Seven LH (one low- and one high-proficient student) or MH (one mid- and one high-proficient student) pairs were identified as the HLPP (4 pairs) or the COP (3 pairs), but none were classified as LPP.

- The homogeneous pairs consisting of mid- and low-proficient students, that is, the MM and LL pairs, did not achieve promising performance. Three of these pairs were identified as COP while the other three were LPP.

- The pairs consisting of one low- and one mid-proficient student (LM) demonstrated diverse levels of performance. One of these pairs was classified as HLPP, two pairs as COP, and the other two as LPP.

Taking together the findings from the cluster analysis and the student artifacts, it seems that the English proficiency differences significantly influenced the students’ pair performance in the collaborative storytelling activities. This result was also supported by the two raters’ observations of the audio recordings of the picture books, which identified even-engagement and uneven-engagement pairs. In 14 even-engagement pairs, the two students took turns to read lines one by one, or they read aloud all the lines together; in seven uneven-engagement pairs, the recording was mainly done by the higher proficient student only. The imbalanced division of labor observed in the uneven-engagement pairs may also have been a result of the proficiency difference in the pair. Due to the critical role of the high proficient students, student pairs including high-proficient students were more likely to be focused on the language quality of their work. On the contrary, student pairs including no high-proficient students tended to shift their attention from linguistic aspects to creativity-oriented performance, and thus were more likely to produce less satisfactory work in terms of its language productivity.

Follow-up interviews

Based on the statistical analysis, interviews were conducted in the hope of explaining the difference in flow perceptions and the use of learning strategies of the high-proficient and low/mid-proficient students. The interviewees were 26 students including seven high-proficient students, 11 mid-proficient students, and eight low-proficient students. All interviews were recorded and then transcribed into text for analysis. The interview
analysis focused on (1) flow-related dimensions (e.g., interest, confidence, challenge, sense of achievement), and (2) language learning strategies (cognitive, memory, metacognitive, compensatory, affective, and social strategies) so as to provide views from different angles than the surveys. Two researchers first identified all texts related to flow perceptions or language learning strategies, then assigned descriptive codes accordingly. Then, the two researchers discussed and merged the codes into themes and categories. Similar codes were merged into themes; for example, codes relating to “fun” and “enjoyment” were merged into the “interest” theme. The two researchers eventually reached consensus on the categorization. After the merging, a total of 33 themes were included in the results, and the numbers of students who mentioned these themes were also counted. In the coding and categorizing process, students’ proficiency levels were also marked so that the interview data could later be interpreted together with the findings from the surveys and artifact analysis. It should be noted that some data could not be coded into the aforementioned categories, such as greetings and students’ making jokes, and so were not included in the analysis.

Several themes regarding students’ flow perception emerged during the interviews, such as interest, confidence, sense of achievement, and challenge. However, these emerging themes were reported by students of different proficiency levels. Firstly, roughly two thirds of the mid- and low-proficient students expressed high interest (fun and enjoyment) and confidence (less fear and no pressure), suggesting that the storytelling activity was fun, enjoyable, and not intimidating; nor was it as stressful as typical English learning activities. However, only two high-proficient students felt the same. One mid-proficient student said, “I became more confident after multiple practices and presentations.” Furthermore, more high-proficient students reported that the storytelling activity was less fun, boring, and not challenging enough for them. Secondly, regarding the sense of achievement in learning, eight mid- and low-proficient students reported a great sense of achievement. One mid-proficient student “began to enjoy the activity later because he became more familiar with the vocabulary,” and a low-proficient student “felt accomplished because she had learned many new words.” Yet again, no high-proficient student expressed a similar sense of achievement. These interview results confirmed the findings from the flow survey (Table 1) that as mid- and low-proficient students felt interested, confident, and perceived a sense of achievement, their flow perceptions increased (Figure 4).

Interview analysis also identified different use of language learning strategies between the mid/low-proficient students and the high-proficient students. First of all, the significant difference lies in students’ use of memory strategies (Table 3). According to the interviews, high-proficient students tended to apply the phonic rules to memorize new words, and the mid- and low-proficient students often used repeating strategies and consulted the e-books for quick answers. Mid- and low-proficient students suggested that they frequently referred to Starfall, the e-book, for the Chinese meaning and learned by “listening to the e-book again and again.” In other words, the mid- and low-proficient students largely relied on the technological features of the e-book, and greatly improved their pronunciation from imitating and repeating after the model recording. In the interviews, eight mid- and low-proficient students reported improvement in pronunciation, while only one high-proficient student mentioned such improvement. Secondly, when mid- and low-proficient students encountered problems in learning English, they usually asked for help from the teachers, parents, and their peers (compensation strategy), which is consistent with the survey findings (Table 3). Lastly, regarding reviewing and planning (metacognitive strategies), some students suggested that they would plan ahead the storyline and review previous work in their picture books; some reported that they would only follow what their partners told them to do. However, use of metacognitive strategies was found across all proficiency levels.

Students’ responses revealed several emerging themes that explain the low flow perception of the high-proficient students. The following interview excerpt is between the interviewer (Interviewer) and a high-proficient student (Student), talking about her collaboration with a low-proficient student.

Interviewer: How do you work with your partner for the digital storytelling activity?
Student: It’s usually like he drew the scenes, and I taught him how to do it.
Interviewer: Do you have any disagreement on this division of labor?
Student: Yes, we do.
Interviewer: Then how do you handle the disagreement?
Student: We would go to the teacher and let her decide.
Interviewer: Why did you disagree with your partner? About what to draw?
Student: ‘Cause I’m good at it.
Interviewer: Do you prefer working alone to working in pairs?
Student: Yes, I’d rather work by myself.
Interviewer: Do you think your partner is helpful?
Student: Well, he is helpful.
Interviewer: But you’d rather work alone?
Student: Yes.
Interviewer: Why is that?
Student: ‘Cause I just don’t like [working with others].

This excerpt displays a typical collaboration pattern between high- and low-proficient students and the imbalanced roles in the student pairs. The higher-proficient students usually played the guiding role to teach their partner English knowledge, and the lower-proficient students received help and contributed to other aspects, such as drawing the scenes for the picture books. Out of the total 26 interviews, 10 students reported similar division of labors, and 11 reported such imbalanced roles. In most pair interactions, while low- and mid-proficient students were able to receive help from high-proficient students, the high-proficient students had to spend time teaching their partners. They took more of a guiding role as they “taught the other student first, and audio recorded the English sentences together when the peer’s reading was okay.” One of the high-proficient students said, “Sometimes, I had to type the exact sentences word by word for him, and teach him how to read every word.”

Discussion

The quantitative and qualitative data were interpreted and discussed together in the hope of providing a clearer understanding of the issue of implementing collaborative storytelling activities in EFL classrooms. It has been indicated in the literature that students’ flow perception in an open-ended learning activity often changes over time (Herrington et al., 2003; Liu et al., 2016; O’Brien & Toms, 2008). Previous studies have also reported that multimedia-rich digital stories capture students’ attention, increase their interest in exploring new ideas, improve language learning motivation, and foster their interest in learning (Di Blas et al., 2009; Figg & McCartney, 2010; Robin, 2008). The results of this study are partially in accordance with the findings of these studies. The mid- and low-proficient students’ flow perceptions increased with time as they gradually gained a sense of achievement in the activity. However, we identified significant interactions between the students’ English proficiency level and their engagement in the activities. High-proficient students’ flow perceptions decreased throughout the activity compared with the low- and mid-proficient students who showed an opposite trend. The decrease in high-proficient students’ flow perception was due to the lack of sense of accomplishment, and the extra burden of playing the guiding role in the pair.

The statistical findings have shown that mid-proficient students, but not high-proficient students, excelled in learning strategy use in this study, which was surprisingly inconsistent with previous literature suggesting that advanced students tend to be more engaged and apply more learning strategies (Uhl Chamot & El-Dinary, 1999; Oxford, 1990). A recent study (Ghani, Mahfuz, & Saad, 2014) also found that advanced learners tended to use more language learning strategies. The different trends in their flow perceptions may explain such a result. As the mid-proficient students perceived higher levels of flow during the latter phases, they were more likely to apply higher levels of learning strategies (Liu et al., 2011; Trevino & Webster, 1992). On the contrary, high-proficiency learners would not apply as many language learning strategies when the challenge level of the task did not closely match their abilities during the later phases of the activity. Therefore, the relevant instruction needs to be dynamically tuned during the learning activity to continuously engage students in flow states from a lower level to a higher level of challenge (Kiili, 2005).

Previous studies have investigated how pair combinations may impact interaction. It was found that when novice students were paired with more capable individuals on a learning task they improved significantly, whilst equal ability pairs did not (Azmitia, 1988; Liu & Tsai, 2008; Rogoff, 1990). In this study, we analyzed how different pair combinations could affect English storytelling performance. This result is in accordance with the finding of previous studies, suggesting that pairs with one student of well-established ability may have favorable performance. This may be because of the centralized knowledge exchange pattern (Liu & Tsai, 2008), as in this study, the high-proficiency students took a guiding role by teaching the other student. Such a centralized interaction pattern led to high-performing pairs because the high-proficiency students were in charge of the progress of the activity. However, such an interaction pattern also caused low-level flow perception among the high-proficiency students as they had to spend extra efforts on teaching and guiding their peers in the activity, rather than completing the challenging learning tasks at their own pace. It was also found that the pairs
consisting of no high-proficiency student did not achieve high level performance in the collaborative storytelling activity. Such a result might be caused by the ability impediment often occurring when no high-achieving student is involved in the collaborative learning groups. Because of the ability impediment, few in-depth meaningful interactions could happen in the collaborative pairs, eventually leading to less satisfactory learning outcomes (Liu & Tsai, 2008). Pedagogical arrangement is necessary to address the needs of the students of different proficiency levels. The finding of a previous study by Rogoff (1990) may provide a solution to address this issue. Rogoff’s study suggested that training a peer to the same level of performance before the collaborative task led to peer dyads performing better than peer dyads in which neither partner had been trained. Therefore, training peers of lower-proficient students to have sufficient skills before the collaborative storytelling activity may also be an effective way to reduce the burden and keep high-proficiency students continuously engaged in the activity. For instance, the teachers may enforce a collaborative script asking students to practice new words of a model story before they take part in the collaborative storytelling activity. In this way, the high-proficiency students’ burden can be reduced.

The collaborative storytelling activity in this present study seemed to limit high-proficient students’ engagement, especially when they were paired with lower-proficient peers. However, it is inevitable to pair or group students of different proficiency levels together for collaborative tasks. Thus, we suggest that, in order to address the pedagogical issues of proficiency differences in one single class, students of different proficiency levels should be guided to play different roles in collaborative activities so as to suit their learning needs and language levels. For example, high-proficient students could be decision-makers when it comes to which English phrases and words to use; lower-proficient students could take part in planning storylines and drawing story scenes. It is suggested that the instructional design of such activities follow the low-threshold-high-ceiling principle (Liu, Cheng, & Huang, 2011; Myers, Hudson, & Pausch, 2000), enabling novice students to participate without difficulties and allowing them to work on increasingly complex products. Meanwhile, the activity design allows a high level of freedom for high-proficient students to contribute and practice more advanced skills, so that they do not lose interest or become disengaged. This low-threshold-high-ceiling design could keep all learners working towards the same shared goal while still allowing different types of participation and accommodating various needs of students of all proficiency levels and enhancing their flow perception at the same time.

Conclusions

According to the findings of this study, students’ engagement, in terms of their flow pattern, strategy use, and pair performance, all changed and varied in accordance with their proficiency levels (Figure 1). Firstly, mid- and low-proficient students’ flow perceptions increased with time as they gradually gained a sense of achievement, while the high-proficient students’ flow perceptions decreased throughout the activity. Secondly, the mid-proficient students were more likely to extensively use diverse categories of strategies in the storytelling activity, especially more than those in the high-proficient group. Thirdly, pair performance largely depended on whether the pair included a high-proficient student who was likely to lead the learning process. Based on these findings, it is suggested that future pedagogical implementations adopt the low-threshold-high-ceiling design so as to address the learning needs of students of different proficiency levels. While activity designs grant easy entry for suffering students and sufficient challenges for high-achievers, they provide opportunities to display different language learning strategies, and help to promote learning from peers in both the linguistic and creative aspects. Along with these educational benefits, students’ engagement and performance might be improved altogether.

It should be noted that this long-term study was mainly based on weekly observation of students’ engagement. The challenge was that such a long-term study required whole classes and the teachers to work with the researchers on a weekly basis for 17 months on top of their regular school schedules. For richer description and deeper analysis of the learning process, we collected and analyzed both quantitative and qualitative data, along with student artifacts to triangulate our findings from the limited sample. Future work could include iterative implementation of the low-threshold-high-ceiling instructional design in different learning contexts with a larger sample size to examine the generalizability and to refine the effective instructional design of collaborative EFL storytelling activities for different age groups.

Acknowledgements

This research was partially funded by the Ministry of Science and Technology, R.O.C. under contract numbers 104-2511-S-008 -014 -MY3 and 103-2511-S-008 -014 -MY3.
References


The Influences of the 2D Image-Based Augmented Reality and Virtual Reality on Student Learning

Hsin-Hun Liou1, Stephen J. H. Yang2*, Sherry Y. Chen2 and Wernhuar Tarng3
1Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan // 2Graduate Institute of Network Learning Technology, National Central University, Taoyuan, Taiwan // 3Institute of Learning Sciences and Technologies, National Tsing Hua University, Hsinchu, Taiwan // 101582004@cc.ncu.edu.tw // jhyang@csie.ncu.edu.tw // sherry@cl.ncu.edu.tw // wtarng@mail.nd.nthu.edu.tw
*Corresponding author

(Submitted January 19, 2016; Accepted June 22, 2016)

ABSTRACT
Virtual reality (VR) learning environments can provide students with concepts of the simulated phenomena, but users are not allowed to interact with real elements. Conversely, augmented reality (AR) learning environments blend real-world environments so AR could enhance the effects of computer simulation and promote students’ realistic experience. However, AR-based learning environments had a lot of dynamic real objects which may increase learners’ mental effort. Moreover, paucity of research compared AR with VR and other mature technologies. Thus, the aim of this study is to compare the influence of the 2D image-based VR and AR in an inquiry-based astronomy course. The findings of this study suggested that the real objects presented in the AR system could reduce the mental load because students could take the real objects of the AR system as the reference objects of the movement of the moon. Furthermore, the sense of the immediacy is increased due to the fact that peers appear on the AR system. Accordingly, the real objects and the sense of the immediacy not only enhance the learning motivations, but also encourage the students to keep conducting the tasks.

Keywords
Elementary education, Interactive learning environments, Media in education, Virtual reality

Introduction

Among a variety of interactive technologies, virtual reality (VR) and augmented reality (AR) share characteristics relevant to simulating a virtual world (Kipper & Rampolla, 2012; Chittaro & Ranon, 2007). According to Milgram, Takemura, Utsumi and Kishino (1994), both AR and VR are contained within the continuum, ranging from a completely real environment to a completely virtual one. VR technology completely replaces a real environment with a dynamic stimulating environment that can be explored interactively by users (Lin & Lan, 2015; Schneps et al., 2014; Wang et al., 2014; Lei, Lin, Wang, & Sun, 2013; Chen, Yang, Shen, & Jeng, 2007; Dori & Barak, 2001), whereas AR refers to technologies that blend real-world environments and context-based digital information (Sommerrauer & Müller, 2014; Azuma, 1997; Milgram et al., 1994).

VR and AR systems integrate different technologies that offer different advantages and disadvantages. Regarding the advantages of VR, only information related to teaching objectives will be considered so VR could provide a simple environment which allows students to test hypothetical scenarios and inquire with a virtual world easily (De Jong, Linn, & Zacharia, 2013; Chen, Yang, Shen, & Jeng, 2007). Moreover, VR could simulate abstract phenomena which facilitate learners to observe and interact with unobservable nature of matter (Wang et al., 2014; Honey & Hilton, 2011; Chen, Yang, Shen, & Jeng, 2007). The benefits of real time visualization could enhance students’ understandings of scientific concepts (Cheng, Lin, & She, 2015; Merchant et al., 2013). Hence, VR has been widely applied in science courses by instructors. The results demonstrated that VR-based learning environments can attract learners’ attention, support inquiry-based learning, and provide students with concepts of the simulated phenomena (Wang et al., 2014; Lee, & Wong, 2014; Chang, Chen, Lin, & Sung, 2008; Chen, Yang, Shen, & Jeng, 2007; Dori & Barak, 2001). Nonetheless, the disadvantage is that users are not allowed to interact with real elements and to have realistic experience. Hence, users cannot easily convert what they learned in the VR system into real situations (Chiu, DeJaegher, & Chao, 2015).

Regarding the advantages of AR, situated information is integrated into a real scene and kinesthetic functions are employed in system design so AR could provide a vivid learning environment in which the effects of computer simulation could be enhanced and students' realistic experience could be promoted. Due to these attractive features, researchers have increasingly recognized AR technologies as a potentially effective method in promoting science learning (Chiang, Yang, & Hwang, 2014; Bressler & Bodzin, 2013; Chang, Wu, & Hsu, 2013; Wu, Lee, Chang, & Liang, 2013; Sollervall, 2012). In particular, researchers have shown that AR-based learning activities not only improve the students’ knowledge construction, but also engage learners in high flow

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.
experience levels (Chiang, Yang, & Hwang, 2014; Sommerauer & Müller, 2014; Ibáñez, Di Serio, Villarán, & Kloos, 2014).

Although AR brought about many positive influences, there are various issues waiting to be explored. Wu et al. (2013) mentioned that comparisons of empirical studies in AR with other more mature technologies may be helpful to highlight the different affordances of AR in learning. Such comparison can help instructors distinguish which scenarios are best suited for AR but not possible with other media. This is the reason why recent studies attempted to compare AR with paper-based, real context or web-based learning materials in the recent decade (Schneers et al., 2014; Zhang et al., 2014; Ibáñez et al., 2014; Martin et al., 2011). For example, a study by Echeverría et al. (2012) compared a multiple mouse game on standard PC and an AR game on tablet computers for co-located collaborative learning. Each student moved their astronaut’s position or changed the setting in a VR scene by controlling an avatar with the mouse wheel in the former game. Conversely, the student moved the real marker in the authentic environment, instead of the avatar to display their astronaut, in the latter game. Specifically, players visualized the augmented world through their Head-Up Display with a mobile platform in the AR game. Their experimental results showed that there were no statistically significant differences in the learning performance between the two groups. In other words, both platforms were effective in increasing the conceptual understanding. Another study by Lin et al. (2013) compared a 2D simulation system with a 3D AR system on the laptop for collaborative learning about an elastic collision. The former enabled users to simulate an elastic collision process in a 2D VR scene while the latter enables the users to manipulate the numerical data and visualize the collision process of the 3D cubes on a real marker. Therefore, the latter one was an AR environment which contained real world objects (Milgram, Takemura, Utsumi & Kishino, 1994). Their results indicated that the AR system not only enhanced the learning achievements, but also enabled dyad learners to respond quickly and support their knowledge construction processes.

The other study by Gavish et al. (2015) evaluated the efficiency and effectiveness of four training groups for industrial maintenance procedural skills. One is the VR platform which displayed the 3D graphics scene to simulate different tasks. Another one is the AR platform which presented the visual information about the current step on real machines. Results demonstrated no significant differences in the final performance between the VR group with the VR platform and the control-VR group with the instructional video. Moreover, the trainees had fewer errors in the AR group with the AR platform than the control-AR group with the real actuator and the instructional film. This study did not directly compare the AR and the VR platform because the two platforms were different in several functions. Therefore, they suggested that studies that compare the effects of VR and AR platforms that were in similar design were being valued.

The aforementioned research demonstrated differences between AR and VR are still uncertain due to inconsistencies results and unsimilar design. Thus, there is a need to evaluate the effects and highlight the different affordances of AR and VR. To this end, this study addresses this issue by investigating the students’ learning achievement, task performance and the acceptance of technology in an astronomy course, where two similar types of observation tools, i.e., the Sky Map and the Moon Finder, are compared. These two observation systems shared similar learning content, but they used different representation methods. The Sky Map was a non-immersive VR product which simulated the celestial bodies with virtual objects while the Moon Finder was an AR software which blended the lunar information and a virtual moon on the real scene. This is the reason why these two observation systems are employed to find answers to the following two research questions:

- Are there any differences in the learning achievement and task performance between a VR-based learning environment and an AR-based learning environment?
- Are there any differences in the acceptance of technology between a VR-based learning environment and an AR-based learning environment?

**Astronomical observation systems**

Two astronomical observation systems were employed in this study. One was the Sky Map while the other was the Moon Finder. Both systems are a simulation of astronomical phenomena on a handheld device and allow users to set different dates and time for displaying the moon. In contrast with simulation, VR system adds the specific requirements of a spatial metaphor (Lee, & Wong, 2014; Chen, Yang, Shen, & Jeng, 2007). These two systems could guide users to find the moon with a kinesthetic interactive method and could provide situated content based on users’ physical movements. Thus, students could simply find the position of the moon with their body, instead of taking complex operations with conventional tools. However, the difference lies within their representation methods.
The VR tools have been developed from two-dimensional (2D) text-based online VR spaces (Lei, Lin, Wang, & Sun, 2013) to three-dimensional (3D) virtual environments (Lin & Lan, 2015; Lee, & Wong, 2014). The Sky Map is a 2d image-based VR product that replicates a real environment with a synthetic virtual world (Kipper & Rampolla, 2012; Chen, Yang, Shen, & Jeng, 2007). The Sky Map simulated virtual celestial bodies with a black sky by specifying location and time (Figure 1). Based on the setting, the synthetic virtual world provided dynamic information, including images and names of the planet, text describing the direction and the grid of the celestial sphere. Moreover, the black sky and the horizon in the Sky Map help users clearly see the shape and the movement of the moon on the screen after searching for the moon so that students cannot be distracted by objects that are irrelevant to learning objectives.

Figure 1. The interface of the Sky Map

<table>
<thead>
<tr>
<th>Date &amp; Time</th>
<th>Direction &amp; Elevation</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013/2/21</td>
<td>15:24</td>
<td></td>
</tr>
</tbody>
</table>

Compass
Lunar phase
Current time
Exit

Figure 2. The interface of the Moon Finder
Conversely, the Moon Finder is an AR software which overlaid digital information in the physical world (Sommerauer & Müller, 2014; Kipper & Rampolla, 2012; Azuma, 1997; Milgram et al., 1994). The Moon Finder showed the image of the moon on the real scene, instead of the black sky (Figure 2). That is, users could see the virtual moon and situated information overlaid on the real scene after finding the moon with their body. By doing so, the degree of reality could be increased because the feature of AR could link the change process of the moon on the daily life environment, instead of a virtual sky. In brief, the Sky Map could support the development of scientific understandings by making students focus on virtual celestial bodies, whereas the Moon Finder could help students link virtual elements to the real-life environment. Thus, different advantages exited in the Moon Finder and the Sky Map due to different representation methods.

On the other hand, the number of virtual objects in the Sky Map was higher than that in the Moon Finder. This is due to the fact that the former simulated planets with virtual objects while the latter integrated dynamic real objects of the physical world. Accordingly, different amounts of multimedia objects, such as an image of the virtual moon, text of the information and animation of the dynamic real objects, were included in the Sky Map and Moon Finder so users might need to use different amounts of working memory and mental effort when interacting with the Moon Finder and the Sky Map. Moreover, Wu et al. (2013) mentioned that students in AR environments may experience cognitive overload because they require to complete complex tasks with multiple technological devices.

Cognitive load theory (CLT) (Chandler & Sweller, 1991) is one of the fundamental theories used to analyze the mental effort and predict the learning effectiveness with new technologies (Chen & Wu, 2015; Sweller, 2010). The CLT proposes that each of three cognitive loads (intrinsic, extraneous and germane) competes for limited resources of working memory and also suggests that the sum of the cognitive load should not exceed the memory resources available (Sweller, 2010). In theory, intrinsic loads are determined by the learning content, extraneous loads do not contribute directly to the understandings of the material, and germane loads help learners process new information and then integrate it into their knowledge structures (Khalil, Paas, Johnson, & Payer, 2005). Appropriate instructional activities and material design can minimize the extraneous load and maximize germane load to ensure effective learning (Sweller, 2010; Khalil et al., 2005; Bodemer, Ploetzner, Feuerlein, & Spada, 2004). In this study, the students may have the similar intrinsic loads because they processed the same complexity of the learning contents. However, they may have different degrees of mental effort, in terms of the extraneous and the germane load, because they were requested to handle different amounts of multimedia objects. Such a difference may cause the students have different levels of extraneous and germane load so that their learning effectiveness might be affected. Therefore, this study is to compare the effects of the Sky Map and the Moon Finder on student learning, in terms of learning achievement and task performance.

Methodology design

An inquiry-based teaching method was applied in curricular activities due to the fact that the observation tools may be helpful in supporting inquiry activities. The Sky Map or the Moon Finder was installed in the tablet PCs, which could provide situational data (date, time, direction and elevation) by combining a digital compass and an accelerometer sensor, which can facilitate students to conduct inquiry activities for astronomy learning. Moreover, a quantitative analysis was employed to compare the learning performance and technology acceptance between VR and AR environments. By doing so, we can investigate the effects and the affordances of VR and AR. The following sections describe the details of the methodology design used in the experiment.

Participants

Two classes of fourth-grade (10 to 11 years old) students in Taiwan took the pre-test. Because the experiment could not easily group the students in random, a quasi-experiment was conducted. Due to not having the parents’ consent to use students’ data, four students were removed from the sample. Moreover, students were also removed if they did not complete the entire post-test. This resulted in an additional two participants being removed from the analysis. In the end, a total of 54 students (Male = 26, Female = 28) participated in this study. Students (all with the same teacher) were separated into teams according to their sciences grades in the last semester for S-type; each team had two or three people and used a tablet PC.
Research instruments

The research instruments adopted in this study were the pre-test, the post-test, the task tests and the questionnaires.

Pre-test and post-test: A “Moon Concept Test” was the pre-test and the post-test which were designed to assess students’ learning performance. The questions in both tests were the same, but appeared in a different order. It consisted of multiple choice items, true or false items, and fill-in-the-blank items, with a perfect score of 100. Both the pre-test and the post-test were taken from textbooks, assignments, teaching guides and prepared materials published by academic textbook publishers, as well as examined by three experts with more than 8 years’ experience of teaching science courses.

Task tests: In order to gather detailed information about the students’ concept of the moon, the task tests were designed to assess students’ understandings of three targeted lunar concepts: the moon phase (Tasks 1), the sequence of the moon phase (Tasks 2), and the moon’s moving trajectory (Tasks 3). Regarding task1, there are eight moon phases in the standard answer. Regarding task2 and task 3, the correct rate was analyzed by circling “yes” or “no” on the sheet.

The Questionnaire: The questionnaire on technology acceptance (Davis, 1989) is beneficial to understand the acceptance of students. The questionnaire used in this study included four aspects: perceived easy to use, usefulness, attitudes, and intention. Each part has four questions, and a 5-point Likert scale was used for all questions, with the scale items ranging from 1 (strongly disagree) to 5 (strongly agree). Participants were required to circle the response that most closely reflected their answer to each question. The reliability of the questionnaire was found to be acceptable (Cronbach’s alpha = 0.89).

Experimental procedures

Figure 3 displays the experimental procedures of this study. The curricular activities took 120 minutes per week and these two groups used the same learning materials, e.g. the book, the worksheet and instructional films. Initially, each student took the pre-test and the initial task test. During the first week, the teacher showed movies and discussed different stories about the moon for students so that old memories were awakened and their learning motivation was stimulated.

![Figure 3. The experimental procedures](image-url)
The next three weeks were a guided inquiry activity that was designed based on the 5E instructional model which is one of the widely-adopted pedagogies in the science learning (Liu, Peng, Wu, & Lin, 2009; Bybee et al., 2006). Five phases are included: engagement, exploration, explanation, elaboration and evaluation. Firstly, the teacher guided students to explore curricular activities through posing inquiry topics and identifying questions on the worksheet (engagement phase). These questions were related to tasks, such as “What are the shapes of the moon during a month?” “Will the moon’s shapes change in a predictable pattern?” “Will the position of the moon be the same at the same time on different days?”, and “What is the shape of the trajectory of the moon during a day?” Next, in order to construct a concrete experiment and build scientific concepts and skills, two groups of students were issued a tablet PC (Figure 4) in which VR and AR observation tools were installed separately, and were tutored in how to collect data related to the moon (exploration phase). Subsequently, the teacher guided students, according to the data that they collected, to discuss and to answer the worksheet questions by writing or drawing. Students also tried to formulate their explanation and the teacher helped students refine their scientific knowledge (explanation phase). Then, the teacher guided each group to use a projector to show their observational results and explain the inferences. Each group shared their opinions and gave feedback to formulate a clear concept (elaboration phase). Thereafter, the participants took the final task test in each week (evaluation phase). Finally, all students took the post-test and fill out the questionnaire. In addition, the researchers interviewed the students to collect their opinions about the interactive experience with the observation systems.

![Figure 4. Students operating the Moon Finder](image)

Results

This study evaluates the effects of two similar design observation systems in an inquiry-based astronomy course, including the Sky Map and the Moon Finder which are associated with VR and AR, respectively. Thus, the experiment included VR group and AR group. The following subsections describe the results on learning achievement, task performance and acceptance of technology.

Learning performance

As shown in Table 1, the independent sample t-test indicated that no significant difference ($F = 0.062, p > .05$) existed in the two groups, in terms of the pre-test scores. In other words, the students had a similar level of prior knowledge. Furthermore, the paired-sample t-test within neither VR nor AR group indicated significant progress ($p < .000$). The results showed that different treatments were equally effective in facilitating desired conceptual change. However, the independent sample t-test indicated that the post-test score of the Moon Finder group was significantly better than that of the Sky Map group ($F = .8.801, p < .05$). As mentioned before, the former used the system with AR while the latter interacted with the VR system. Accordingly, it seemed that the AR group demonstrated better learning performance than the VR group.
Regardless of the AR group or VR group, the participants needed to conduct three tasks. Table 2 describes the students’ task performance of these tasks. Regarding Task 1 (Scientific moon phase), no significant differences exist between the AR group and VR group, in terms of their initial performance. Likewise, the AR group and VR group also demonstrated similar performance, in terms of their final performance ($F = 3.322$, $p = 0.07 > .05$). The results suggested that both of AR and VR are efficient for students to learn this subject.

**Table 1. The independent sample t-test for pre-test and post-test across different observation systems**

<table>
<thead>
<tr>
<th></th>
<th>VR</th>
<th>AR</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>21.15(8.075)</td>
<td>22.15(8.782)</td>
<td>0.062</td>
<td>0.665</td>
</tr>
<tr>
<td>Post-test</td>
<td>64.76(16.76)</td>
<td>73.55(10.78)</td>
<td>8.801</td>
<td>0.028*</td>
</tr>
</tbody>
</table>

Note. *$p < .05$.  

Regarding Task 2 (Scientific moon sequence), both groups demonstrated similar initial performance. After interacting with the system assigned to them, these two groups significantly performed differently. More specifically, the AR group significantly performed better than the VR group ($F = 11.431$, $p < .05$). Regarding Task 3 (The trajectory of the moon), participants in the AR group initially performed similarly to those in the VR group. However, the AR group significantly performed better than the VR group for their final performance ($F = 0.33$, $p < .05$). In brief, the AR group significantly performed better than the VR group in Task 2 and Task 3, not but Task 1.

**Table 2. Participants’ responses coded as scientific**

<table>
<thead>
<tr>
<th>Question</th>
<th>VR</th>
<th>AR</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task 1: Scientific moon phase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Performance</td>
<td>2.15(1.350)</td>
<td>1.93(0.874)</td>
<td>3.041</td>
<td>0.476</td>
</tr>
<tr>
<td>Final Performance</td>
<td>6.04(1.755)</td>
<td>6.81(1.272)</td>
<td>3.322</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>Task 2: Scientific moon sequence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Performance (%)</td>
<td>26 (.447)</td>
<td>30 (.465)</td>
<td>0.356</td>
<td>0.767</td>
</tr>
<tr>
<td>Final Performance (%)</td>
<td>48 (.590)</td>
<td>78 (.424)</td>
<td>11.431</td>
<td>0.024*</td>
</tr>
<tr>
<td><strong>Task 3: The trajectory of the moon</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Performance (%)</td>
<td>19 (.396)</td>
<td>15 (.362)</td>
<td>0.518</td>
<td>0.721</td>
</tr>
<tr>
<td>Final Performance (%)</td>
<td>30 (.465)</td>
<td>67 (.480)</td>
<td>0.33</td>
<td>0.006*</td>
</tr>
</tbody>
</table>

Note. $p < .05$.  

Acceptance of technology

As mentioned previously, the questionnaire covered four aspects, which were applied to identify learners’ acceptance of VR or AR systems. Among these four aspects, no significant differences were found for “Easy for use” and “Intention.” On the other hand, significant differences were found for “Usefulness” and “Attitude.” Table 3 presents the mean value and standard deviation of each statement of these two aspects. Regarding usefulness, the AR group agreed that the tool could not only help them understand the sequence of the moon phase easily (Q6 ($F = 24.032$, $p < .05$), but also could help them identify the moving direction of the moon easily (Q7 ($F = 14.542$, $p < .000$). Additionally, they tended to agree that using the tool to observe the moon was exciting (Q11 ($F = 0.022$, $p < .05$).

**Table 3. Means and standard deviations for the acceptance**

<table>
<thead>
<tr>
<th>Question</th>
<th>VR</th>
<th>AR</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Usefulness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q06: I can understand the sequence of the moon phase easily with the tool.</td>
<td>3.33(1.641)</td>
<td>4.26(0.813)</td>
<td>24.032</td>
<td>0.011*</td>
</tr>
<tr>
<td>Q07: I can understand the moving direction of the moon easily with the tool.</td>
<td>3.11(0.577)</td>
<td>4.30(0.953)</td>
<td>14.542</td>
<td>0.000**</td>
</tr>
<tr>
<td><strong>Attitude</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11: Using the tool to observe the moon is excited.</td>
<td>3.30(1.103)</td>
<td>4.00(1.144)</td>
<td>0.022</td>
<td>0.025*</td>
</tr>
</tbody>
</table>

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

These results suggested that the AR group tended to have positive perceptions that may influence the task performance. These results echo those presented in the results of the learning performance, which indicated that students in the AR group had better correct rate than the VR group on Task 2 and Task 3. Conversely, the VR group tended to have negative perceptions. For example, they disagreed that the system’s functions could help...
them understand the sequence of the moon phase and the moving direction of the moon easily. Moreover, they were not so excited when using the tool. In brief, the AR group showed more positive perceptions than the VR group, in terms of the “Usefulness” and “Attitude.”

Qualitative data

In addition to the quantitative results presented in the previous two sections, the observed data and interview results were analyzed. After analyzing learners’ responses to these two systems, we realized that learners in the AR group perceived the usefulness because they could use the objects in the real environment as a reference to describe the position of the moon, in terms of the direction and elevation. Conversely, learners in the VR group experienced less usefulness because they could not easily locate the position of the moon with the real objects in the environment. The details of learners’ responses are described in Table 4.

The responses shown in Table 4 can further be employed to identify why augmented reality has such different impacts on these two groups. More specifically, the AR group used the objects in the real environment as a reference in determining the moon’s position after students located the moon. Furthermore, they were glad to talk about the shape of the moon and used the location of his/her classmate as a reference. However, the VR group needed to image the virtual scenario in the real environment. They were less excited because they could not easily immerse in the environment. Accordingly, the students in the VR group felt it uneasy to remember the moving direction of the moon. In brief, these two groups demonstrated different responses during the experimental process. Most of the students argued that the real objects in the AR group were not only useful for students to describe the position of the moon, but also made them to have positive perceptions.

Table 4. Students’ responses to the different systems

<table>
<thead>
<tr>
<th>The VR group (N = 15)</th>
<th>The AR group (N = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>After I adjusted the time, the moon has gone up to a different position. Let us record the data and then draw its trajectory.</td>
<td>Initially, the moon’s position was above the blackboard. After I adjusted the time, the moon had gone up to the ceiling. I found that its direction and elevation were changed.</td>
</tr>
<tr>
<td>I followed the arrow to find the moon. Its shape was probably changed. Let’s check the record data.</td>
<td>I can see you on the screen with the moon. Moreover, it is interesting to see that the moon’s shape just looked like the half of your right face.</td>
</tr>
<tr>
<td>I found the moon rise from the east and down in the west.</td>
<td>I found the moon rise from the east and the direction was the same with the school door.</td>
</tr>
</tbody>
</table>

Discussions

Astronomy is an abstract scientific concept and is cognitively demand, especially for elementary school students. In this study, a quasi-experiment was conducted to evaluate the effects of VR and AR environments on an astronomy inquiry course. The results showed that both technologies had significant effects on learning performance. Students in the both groups agreed that they not only felt the systems were easily operated but also had the intention of using the systems. However, statistical differences exist between the VR and AR in terms of the performance and the acceptance (Figure 5).

Such a difference might be caused by the fact that the mental resources used in the learning activities do not contribute directly to constructing the schema. More specifically, the Sky Map system used by the VR group stimulated the virtual celestial body on the screen, and was an isolated virtual environment. Thus, students had to put greater mental effort toward exploring the rules of the direction and shape of the moon. On the other hand, the AR system, which integrated virtual objects and the real environment, allowed students to easily connect the moon’s position and shapes with reference to the real world. Therefore, the information about the date, time, direction, elevation, and shape of the moon and real objects were shown on the same screen. This may lead to decrease mental load evoked by the instructional material. In brief, AR and VR may lead students to bear different degrees of mental load, and this difference also caused students perform differently and to react differently to the usefulness of these two systems.
Additionally, the fact that the students in the AR group performed better in the last two tasks may be related to the perception of immediacy. Previous studies pointed out that immediacy is the students’ sense of a realistic context which not only gives learners a sense of being in a place with others but also can achieve high-quality interaction with the learning environment (Kotranza, Lind, Pugh, & Lok, 2009). Because students in the AR group saw their classmates on the screen, their sense of immediacy may be increased and their concentration may also be improved. Moreover, students in the AR group tended to have positive emotions which could make them engage in the learning activities so their task performance could be improved. On the other hand, real objects were not connected to the instructional material in the VR environment and classmates were not shown on the screens of the devices. Accordingly, students in the VR group had a lower-quality interaction with the real environment and had more serious learning experience. Therefore, students’ concentration and their sense of immediacy might have been low. In brief, the sense of immediacy existing in the two groups had great effects on students’ task achievement and attitude.

In general, these findings are in line with the results by Lin et al. (2013), which indicated that an AR simulation system could affect learning performances. Specifically, Lin et al. (2013) pointed out that the AR system may facilitate students responding quickly to the displayed results, which could increase students’ knowledge construction process. Furthermore, this study shows that the AR-based environment may increase the students’ perception of immediacy and enjoyment, thereby promoting students’ concentration. In brief, the AR system could make learners integrate multimedia elements with positive emotions so that the mental load could be reduced.

**Conclusions**

Although AR technology has been proven positive in science learning, little research has compared AR with other more mature technologies. The aim of this study was to compare the effectiveness of the VR and AR systems in an inquiry-based astronomy course. Regarding the learning performance and task performance, the findings from this study revealed that the students in the AR group performed significantly better than those in the VR group. Accordingly, these findings imply that the AR technology is helpful for students to learn a moon phase course. Regarding the acceptance, “Easy for use” and “Intention” of the acceptance between the two groups did not show significant differences, significant differences were found for “Usefulness” and “Attitude.” With the features of the AR system, learners can easily integrate virtual objects and real environments so as to decrease the mental load and improve their learning. For example, they may connect the information from the media with the real objects and understand the rules of the phenomena. Moreover, the sense of immediacy in the AR group may be higher than those in the VR group. This may improve the students’ positive learning experience and concentration in the learning process.

This study makes a contribution mainly on two aspects: theories and applications. Regarding the theories, the findings of this study indicated that the real objects in the AR system could reduce the mental load of the CLT and increase the immediacy because students could take the real objects of the AR system as the reference.
objects of the movement of the moon. Moreover, peers appear on the system so the sense of immediacy is increased. Accordingly, these factors not only enhance the learning motivations, but also encourage the students to keep conducting the tasks. Such findings deepen the understandings of the effectiveness of the AR system by providing empirical evidence.

In terms of the applications, the study describes how to implement the AR system or the VR system in inquiry activities and provides evidence that both systems could improve the students’ knowledge construction. In other words, both AR and VR systems are beneficial for knowledge construction. Thus, the approaches used to implement the AR system or the VR system in this study can be used to guide instructional designers how to implement AR and VR systems.

The experimental results show that the AR-based environment was beneficial for improving learning achievement and task performance. Nevertheless, this study has several limitations. Firstly, our results were obtained by using a non-immersive VR environment; as such, a truly immersive VR system could be used to fully compare the students’ experience with an AR environment in the future. Secondly, due to the relatively less qualitative results to support the quantitative results, more qualitative results should be addressed in future work, such as behavior pattern analysis and eye-tracking techniques. As recommended by Cheng and Tsai (2013), different personal characteristics or learning status still need to be investigated when involved in AR systems. Therefore, researchers are encouraged to explore the relationships between learning styles and the use of an AR-based educational environment. Such findings could provide guidance how to implement adaptive AR learning systems and make the learning process more effective.

Acknowledgments

Funding for this research work is provided by the Ministry of Science and Technology, Taiwan, under Grant Nos. NSC 101-2511-S-134-002, 102-2511-S-134-009, NSC 102-2511-S-008-013-MY3, and MOST 104-2511-S-008-006-MY2.

References


An Investigation of Technological Pedagogical Content Knowledge, Self-Confidence, and Perception of Pre-Service Middle School Mathematics Teachers towards Instructional Technologies

Ilhan Karatas*, Mutlu Piskin Tunc, Nurbanu Yılmaz and Gulzade Karaci
Bulent Ecevit University, Faculty of Education, Kdz Eregli, Zonguldak, Turkey // ilhankaratas@beun.edu.tr // mutluptunc@beun.edu.tr // nurbanuyilmaz@beun.edu.tr // gulzade.karaci@ beun.edu.tr
*Corresponding author

(Submitted January 27, 2016; Revised June 30, 2016; Accepted July 21, 2016)

ABSTRACT
Technology provides new methods and approaches for educational activities. Therefore, teachers should improve their ability and knowledge to integrate technology into instruction. The use of technology-based learning environment which is effectively used to improve the technological pedagogical content knowledge of pre-service teachers has a crucial importance for the training of pre-service teachers. In this regard, the purpose of this study is to investigate the technological pedagogical content knowledge (TPACK), TPACK related self-confidence, and perception of pre-service middle school mathematics teachers in terms of instructional technologies. In this study, TPACK Survey, TPACK Self-Confidence Survey, and TPACK Perception Survey were administered to 427 pre-service middle school mathematics teachers in elementary mathematics education program. The data were analyzed quantitatively. The data analysis revealed that there was a significant relationship between gender and perception towards technology. Moreover, it might be concluded that pre-service teachers improve their knowledge and self-confidence to use technology in elementary mathematics education programs. Lastly, considering the finding that there was a relationship between the use of technology and self-confidence towards the use of technology, it might be inferred that self-confidence of pre-service teachers towards the use of educational technologies increases with the use instructional tools.

Keywords
Technological pedagogical content knowledge, Self-confidence, Perception, Pre-service teachers

Introduction
Teachers are individuals who provide appropriate methods, techniques, and materials in the process of education either as the source of the information or as a guide during the teaching process (Yalın, 2000; Şimşek, 2000). In today’s world, technology has become an important part of teachers’ knowledge base. Therefore, this makes the use of technology in education more frequent (Graham, Burgoyne, Cantrell, Smith, Clair, & Harris, 2009; Öztürk & Horzum, 2011).

Since the integration of technology has become crucial in education, the knowledge of teachers related to the use of technology has been centered at the heart of the research on educational technology. As it is defined previously, technological pedagogical content knowledge (TPACK) is the knowledge acquired by teachers in order to integrate technology in education while teaching content to the students (Schmidt, Baran, Thompson, Mishra, Koehler, & Shin, 2009). Although several studies have been conducted in the scope of TPACK, there is no obvious form of the technological knowledge that teachers should acquire and a strict method on how it should be acquired by teachers (Koehler & Mishra, 2005). Merely introducing the technology in educational process does not create a significant change on the integration of technology in education (Carr, Jonassen, Marra, & Litzinger, 1998). Therefore, how the technology can be integrated in education should be investigated in more detail since the knowledge of technology cannot be separated from the knowledge of teaching context in educational technology (Koehler & Mishra, 2005). In addition, it is recognized that the use of technology alone cannot create an effective improvement on the learning of students. Graham et al. (2009) emphasized that just the use of technology cannot get students to learn effectively. Rather, teachers should know how to use the instructional technology during the teaching process. Similarly, Clark (1985) highlighted the difference between “media” and “method of instruction.” That is, mere the use of media cannot provide learning. It is a tool providing learning by the employment of appropriate methods and techniques while teaching mathematics. Therefore, the effective use of instructional technologies is more important than just acquisition of them for educational purposes.

Although technology has become available in the classrooms, the use of it has continued to be criticized because of some factors, such as teachers’ use of technology infrequently and for knowledge transmission rather than the
construction of knowledge (Clark, 1985; Gao, Choy, Wong, & Wu, 2009; Harris, Mishra, & Koehler, 2009; Ottenbreit-Leftwich, Glazewski, Newby, & Ertmer, 2010; Sang, Valcke, van Braak, & Tondeur, 2010). Therefore, those observations get the correct integration of technology more emphasized in education (Chai, Koh, Tsai, & Tan, 2011).

In Turkey, FATIH Project (The Project of Improving Opportunities and Instructional Technologies) has been conducted since 2010 in order to integrate technology into education across the country (The Office of Innovation and Educational Technologies, 2016). One of the key actors of this project is teachers. Therefore, ideas and competences of teachers toward technology integrated education is an important factor for the effectiveness of the project. For this reason, it is important to investigate the integration of technology in education from the perspectives of teachers. As researchers stated, there is a need to develop teacher education strategies in terms of teachers’ effective integration of technology into their teaching (Koehler & Mishra, 2005). Thus, an effective professional teacher education program can be prepared in order to get teachers be oriented with instructional technologies (Karataş, 2014a). In order to develop strategies to integrate technology in education, it is crucial to investigate the views of teachers based on the integration of technology into teaching (Öksüz, Ak, & Uça, 2009). For this reason, ideas and competences of pre-service teachers in relation with instructional technologies were investigated within the scope of the current study. In addition, the investigation of views of pre-service teachers makes a great contribution to the development of current teacher education program (Tınmaz, 2004). As Bitner and Bitner (2002) stated, there are several factors which are important to get teachers to integrate technology into teaching successfully. One of the factors is labeled as fear of change which includes self-confidence and perception. As Christensen (1997) stated, self-confidence levels of teachers in relation with technology use affect their teaching in learning environment. Also, teachers with favorable perception of technology are more eager to use technology in education (Tınmaz, 2004). Therefore, beliefs and ideas of teachers related to the use of technology might provide a valuable insight for researchers. As Tınmaz (2004) suggested gender is one of the major factors affecting the perception and competency of pre-service teachers in relation with technology use in education. Therefore, gender is one of the elements investigated within the current study. Moreover, TPACK, TPACK related self-confidence, and TPACK related perception of pre-service teachers were investigated based on grade level of pre-service middle school mathematics teachers in order to analyze whether there is a difference or not among grade levels. As Dong, Chai, Sang, Koh, and Tsai (2015) proposed for Chinese teachers, the efficiency of teacher education programs can be improved by a better understanding the TPACK and TPACK related beliefs of teachers. Moreover, Paraskeva, Bouta, and Papagianni (2008) stated that perception and self-confidence of teachers can be changed by the use of technological tools in teacher education. For this reason, the TPACK related content knowledge, perception, and self-confidence of pre-service teachers were investigated in the current study. This study is expected to contribute to the literature and close the gap in order to overcome the complexity of the integration of technology into teaching environments.

**Technological pedagogical content knowledge**

Technological pedagogical content knowledge (TPACK) is the knowledge acquired by teachers in order to integrate technology into education while teaching a particular content to students (Schmidt, Baran, Thompson, Mishra, Koehler, & Shin, 2009). Although several studies have been conducted on TPACK, there is no explicit form of the technological knowledge that teachers should acquire and a strict method on how it should be acquired by teachers (Koehler & Mishra, 2005). Therefore, further studies are required in order to examine the effects of TPACK on teachers’ use of instructional technologies. In addition to the TPACK, perception of teachers toward instructional technologies is the other factor affecting teachers’ use of instructional tools which will be described briefly.

**Perceptions towards instructional technologies**

According to Teo (2010), teacher educators should study beliefs and perceptions of teachers in order to grasp those characteristics of teachers and prepare courses for effective integration of technology. “Can pre-service teachers move beyond their own perceptions to create new visions of what teachers can be in the future?” or “Will developing a critical consciousness help teachers create these new visions of education?” are some questions directed by researchers related to perceptions of teachers (Carr-Chellman & Dyer, 2000, p. 4). Some researchers answered those questions by stating that teachers often teach the way they were taught in the past (Carr-Chellman & Dyer, 2000). Therefore, it might be relatively challenging to get teachers to use instructional technologies if they were not taught with those methods. According to Teo (2010), it is reasonable to create and
maintain a positive attitude toward learning technologies in the process of pre-service teacher education rather than after they begin the profession. Moreover, the researcher stated that attitude of teachers can be examined in order to predict the future use of technological tools by pre-service teachers. Therefore, the investigation of attitude of pre-service teachers towards learning technologies may give clues about their professional career. According to the research of Yıldırım (2000), the teachers who used computers more tend to develop positive attitudes which promote further use of learning technologies in teaching tasks such as online forums and ICT-based student-centered learning. Moreover, the perception of teachers towards learning technologies is a factor which has an influence on the successful integration of technological devices into teaching (Parr, 1999). There have been various studies in order to understand the perceptions of teachers on instructional technology (Savenye, 1992; Downes, 1993; Diegnueller, 1992). In the scope of those studies, whether teachers will use instructional technologies in their future career, their feelings about their competence, the notion of the relationship of technology and change in their minds, and how often technological tools are utilized in their methods courses were investigated.

The other concern is the perceptions of students since students’ perceptions tend to be influenced by the perceptions of teachers and also by the use of technological devices within teaching and learning process as it is presented in Figure 1 (Cope & Ward, 2002). As the study of Cope and Ward (2002) illustrated, inadequate knowledge and improper perceptions of teachers towards instructional technologies may hinder the successful integration of those technologies and their respective learning outcomes. Therefore, it can be inferred that the attitude of teachers towards computers is a central concern to achieve successful use of computer in the teaching process (Teo, 2010). As a result, studies on teacher perception are important because of its effects on perceptions and learning of students in the long run.

While performing a task, self-efficacy is also a factor affecting one’s preferences (Bandura, 1977). Therefore, self-efficacy beliefs of teachers might affect the preferences of teachers in relation with the use of instructional technologies in education. Therefore, self-confidence as a measure of self-efficacy will be briefly explained (Bandura, 1986).

**Self-confidence towards instructional technologies**

Literature shows that teachers tend to use more technology in their classroom if they have high levels of confidence in instructional technologies (Dawson, 2008; Zhao, Pugh, Sheldon & Byers, 2002; Littrell, Zagumny & Zagumny, 2005). According to the study of Graham et al. (2009), confidence in technological knowledge is a prerequisite for the confidence in TPACK. In other words, basic technical knowledge and skills related to instructional technologies are necessary in order to integrate them into instruction effectively (Finger, Jamieson-Proctor, & Albion, 2013). Confidence is accepted as a measure of self-efficacy (Bandura, 1986; Shell, Murphy, & Bruning, 1989). Therefore, the confidence level of pre-service teachers towards the use of instructional technologies is investigated as a measure of self-efficacy within the current study. Self-efficacy is one’s belief related to his/her capacity to perform a particular task (Bandura, 1977). Therefore, computer self-efficacy can be defined as one’s belief related to the use of computers for a specific purpose (Compeau & Higgins, 1995). Since inappropriate belief on computers results with less likely use of them, the self-efficacy level of teachers can be used to predict teachers’ integration of technology within the teaching process (Delcourt & Kinzie, 1993; Oliver & Shapiro, 1993). According to Pamuk and Peker (2009), gender is the variable which is mostly studied in
relation with computer self-efficacy in comparison with other variables such as age, socio-economic status, and computer experience in the literature. According to the results, there has been no agreement in the literature on the relationship between computer self-efficacy and gender. In some studies males performed a higher computer self-efficacy (Dunndell & Haag, 2002; Işıkşal & Aşkar, 2003) whereas in some other studies there was no significant difference between male and female participants (Akkyunlu & Orhan, 2003). Moreover, it is found that there is a correlation between computer self-efficacy and experience and training on the use of computers (Marakas, Yi & Johnson, 1998; Wilfong, 2006). Therefore, the training of teachers related to the use of learning technologies is significant in order to equip teachers with the required skills, attitudes, and knowledge toward learning technologies (Pamuk & Peker, 2009).

As researchers stated, there is a need to develop teacher education strategies in terms of effective integration of technology into teaching (Chai, Koh, Tsai, & Tan, 2011; Koehler & Mishra, 2005). In order to develop strategies to integrate technology into education, it is crucial to investigate the belief and ideas of teachers related to integration of technology into teaching (Öksüz, Ak, & Uça, 2009). Therefore, the knowledge and views of teachers related to TPACK might provide a valuable insight so as to get teachers involve the technology into their teaching effectively. For this reason, the TPACK related content knowledge, perception, and self-confidence of teachers were investigated within the current study.

The purpose of the study

The purpose of the study is to investigate the TPACK, TPACK related self-confidence, and perception of pre-service middle school mathematics teachers towards technology use.

Method

Study context

Faculties of education are the primary teacher education institutions in Turkey. Middle school mathematics teachers (5th through 8th) have a four-year undergraduate education in elementary mathematics education program while secondary level (9th through 12th) mathematics teachers are certified by 5-year long secondary mathematics education program. Both elementary and secondary mathematics education programs include content and pedagogical content courses specific to the needs of each grade level. Some of the content courses and all of the pedagogical content courses offered in both programs are similar in nature. However, because of the different content domains between elementary and secondary levels, secondary mathematics teacher education programs have advanced level content courses different from elementary mathematics teacher education programs.

The current teacher education program includes the courses comprising ICT (Instructional Computer Technologies) in an effort to train teachers to teach by the use of technology. There are two courses related to instructional technologies taught over two semesters in the first year of the program. Those courses include basic information technologies such as word processing, spreadsheets and presentation software to pre-service teachers. Those courses have been offered at almost each education faculty for four hours a week over two terms in Turkey. Teacher educators may prefer to use software such as PowerPoint in their undergraduate courses which can be considered as a starting point for the use of instructional technologies (Karataş, 2014b). However, the instructional experiences of pre-service teachers with technology do not go beyond the presentations given by the lecturers. Consequently, it is useful to examine the beliefs and knowledge of pre-service teachers concerning the use of technology for mathematics education. Since those courses cannot provide real teaching experiences with technology to pre-service teachers, some education faculties may also offer elective courses such as “Using Technology in Teaching and Learning Mathematics.” Those courses present theoretical bases for computer-based mathematics education and introduce some commonly used mathematical software, such as Dynamic Geometry Systems, Computer Algebra Systems and Win Logo. Within the scope of this course, pre-service teachers may also be assigned with mini-projects. The fourth grade pre-service middle school mathematics teachers also took the course “Using Technology in Teaching and Learning Mathematics” before the current study. Moreover, third and fourth grade pre-service elementary mathematics teachers take the course “Methods of Teaching Mathematics I-II.” In this course, pre-service mathematics teachers learn methods of teaching mathematics by the use of particular methods and techniques. Those courses might affect TPACK, TPACK related self-confidence, and TPACK related perception of students.
Participants

The participants of the study were pre-service middle school mathematics teachers who had adequate knowledge to use the technology in mathematics education. The data collection tools were administered to 427 pre-service middle school mathematics teachers, 104 freshman, 77 sophomore, 193 junior, and 53 senior students in elementary mathematics education program.

Data collection

In order to explore TPACK, TPACK related self-confidence, perception of pre-service teachers on the use of technology, and the relationship among them, three surveys were administered to pre-service teachers. In the following subtopics, the data collection tools are described briefly.

Technological Pedagogical Content Knowledge Survey (TPACK-S)

The survey was used in order to examine the TPACK of pre-service teachers. It was developed by Schmidt et al. (2009) and translated into Turkish by Öztürk and Horzum (2011). The survey was a 5-point Likert survey composed of 47 items across 7 dimensions, technological knowledge, content knowledge, pedagogical knowledge, pedagogical content knowledge, technological content knowledge, technological pedagogical knowledge, and technological pedagogical content knowledge. The five points of Likert survey were “Totally agree,” “Agree,” “Neutral,” “Disagree,” and “Totally disagree.” In order to examine the construct validity of the scale, exploratory and confirmatory factor analyses were conducted by Öztürk and Horzum (2011). As a result, the scale was found similar to the original scale. The reliability of the survey was calculated as 0.94 by the use of Cronbach alpha coefficient.

Technological Pedagogical Content Knowledge Self-confidence Survey (TPACK-SCS)

The survey was used in order to examine TPACK related self-confidence of pre-service teachers. It was developed by Graham et al. (2009) and translated into Turkish by Timur and Taşar (2011). The survey was comprised of 31 items including technological pedagogical content knowledge, technological pedagogical knowledge, technological content knowledge, and technological knowledge dimensions. The Cronbach alpha reliability coefficient was calculated as 0.92 by Timur and Taşar (2011). Since the Cronbach alpha coefficient of a scale should be above .7 (Pallant, 2005), the value might be considered reasonable for this study. The result of the factor analysis, which was conducted to determine the construct validity of the scale, showed that the scale was also valid in Turkish context and its structure was acceptable (Timur & Taşar, 2011).

The survey of perception towards technology

The survey was used in order to examine the perception of pre-service teachers regarding the use of technology in mathematics education. The survey developed by Öksüz, Ak, and Uça (2009) was comprised of 73 items. The internal consistency among the items was calculated as 0.95 by the use of Cronbach Alpha coefficient. As Pallant (2005) stated, Cronbach alpha coefficient of a scale should be greater than .7. Therefore, the value 0.95 was considered reasonable for the current study.

Data analysis

TPACK, TPACK related self-confidence, and perceptions of pre-service teachers regarding the use of technology in education were examined by the use of three different surveys. One-way between-groups multivariate analysis of variance (MANOVA) test was used in order to investigate whether there was a statistically significant difference between groups of male and female pre-service teachers. Then, another MANOVA test was conducted to observe whether there was a significant difference among the grade levels of pre-service teachers. Bonferroni adjustment was used to control for the Type 1 error. To do this, normal alpha value (.05) was divided by the number of tests that were used. Since there were three surveys, .05 was divided by 3 (which equal to .017 after rounding) and this new value was used to determine the significance (Pallant, 2005).
Results

The means and standard deviations were calculated by the use of points gathered from the surveys in order to examine the TPACK, TPACK related self-confidence, and perceptions of pre-service teachers regarding the use of technology in terms of gender (See Table 1). Table 1 shows that the mean of the male pre-service teachers on technological pedagogical content knowledge survey (TPACK-S) ($M = 3.44$, $SD = .43$) and technological pedagogical content knowledge related self-confidence survey (TPACK-SCS) ($M = 3.59$, $SD = .57$) was close to, but higher than the mean of the female pre-service teachers on TPACK-S ($M = 3.38$, $SD = .45$) and TPACK-SCS ($M = 3.46$, $SD = .56$). However, it was observed that female pre-service teachers ($M = 3.74$, $SD = .42$) performed better than male pre-service teachers ($M = 3.69$, $SD = .44$) on the Perception Survey related to use of technology.

<table>
<thead>
<tr>
<th>Survey</th>
<th>Gender</th>
<th>$N$</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPACK-S</td>
<td>Female</td>
<td>330</td>
<td>3.38</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>97</td>
<td>3.44</td>
<td>.43</td>
</tr>
<tr>
<td>TPACK-SCS**</td>
<td>Female</td>
<td>330</td>
<td>3.46</td>
<td>.56</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>97</td>
<td>3.59</td>
<td>.57</td>
</tr>
<tr>
<td>Perception</td>
<td>Female</td>
<td>330</td>
<td>3.74</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>97</td>
<td>3.69</td>
<td>.44</td>
</tr>
</tbody>
</table>

Note. $M =$ Mean; $SD =$ Standard deviation. *TPACK-S = Technological Pedagogical Content Knowledge Survey; **TPACK-SCS = Technological Pedagogical Content Knowledge Self Confidence Survey.

One-way between-groups multivariate analysis of variance (MANOVA) was used to investigate gender differences between scores gathered from the surveys. Preliminary assumption testing was conducted to check the normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Type III sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPACK-S</td>
<td>.278</td>
<td>1</td>
<td>.278</td>
<td>1.400</td>
<td>.237</td>
<td>.003</td>
</tr>
<tr>
<td>TPACK-SCS**</td>
<td>1.402</td>
<td>1</td>
<td>1.402</td>
<td>4.451</td>
<td>.269</td>
<td>.003</td>
</tr>
<tr>
<td>Perception</td>
<td>.219</td>
<td>1</td>
<td>.219</td>
<td>1.225</td>
<td>.035</td>
<td>.010</td>
</tr>
</tbody>
</table>

Note. *TPACK-S = Technological Pedagogical Content Knowledge Survey; **TPACK-SCS = Technological Pedagogical Content Knowledge Self Confidence Survey.

There was a statistically significant difference between males and females on the combined dependent variables: $F(3,423) = 2.95$, $p = .032$; Wilks’s Lambda = .98; partial eta squared ($\eta^2$) = .02. If the results for the dependent variables were considered separately, Table 2 revealed there was no statistical significance using a Bonferroni adjusted alpha level of .017 (0.05 divided by 3 based on the number of dependent variables), for any variable.
Partial eta squared (effect size) ranged from .010 to .003, each of these is a relatively small effect size, according to Cohen (1988).

The means and standard deviations were calculated by the use of points gathered from the surveys in order to examine the TPACK, TPACK related self-confidence, and perception of pre-service teachers regarding the use of technology based on grade levels of the pre-service teachers. Table 3 illustrates that freshman (1st grade) pre-service teachers had the lowest mean scores while the senior (4th grade) pre-service teachers had the highest mean scores on each of the surveys. Moreover, sophomore (2nd grade) pre-service teachers had close to but higher than the junior (3rd grade) pre-service teachers on the surveys. One-way between-groups multivariate analysis of variance (MANOVA) was used to investigate grade level differences in scores obtained from the surveys. Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity, with no serious violations noted.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Type III sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPACK-S</td>
<td>1.571</td>
<td>3</td>
<td>.524</td>
<td>2.668</td>
<td>.047</td>
<td>.019</td>
</tr>
<tr>
<td>TPACK-SCS**</td>
<td>6.379</td>
<td>3</td>
<td>2.126</td>
<td>6.980</td>
<td>.000</td>
<td>.047</td>
</tr>
<tr>
<td>perception</td>
<td>1.429</td>
<td>3</td>
<td>.476</td>
<td>2.697</td>
<td>.046</td>
<td>.019</td>
</tr>
</tbody>
</table>

*Note. TPACK-S = Technological Pedagogical Content Knowledge Survey; TPACK-SCS = Technological Pedagogical Content Knowledge Self Confidence Survey.

As Table 4 shows, there was a statistically significant difference between grade levels on the combined dependent variables: \(F(9,1269) = 2.81, p = .003\); Wilks’s Lambda = .94; partial eta squared(\(\eta^2\)) = .02. If the results for the dependent variables were considered separately, the only difference to reach statistical significance using a Bonferroni adjusted alpha level of .017, was TPACK-SCS: \(F(3,423) = 6.980, p = .000\), partial eta squared(\(\eta^2\)) = .047. Post-hoc comparisons using the Tukey’s HSD test indicated that the mean score for Grade 1 (\(M = 3.32, SD = .65\)) was significantly different from Grade 2 (\(M = 3.55, SD = .53\)) and Grade 4 (\(M = 3.73, SD = .48\)). In addition, Grade 3 (\(M = 3.49, SD = .52\)) was significantly different from Grade 4 (\(M = 3.73, SD = .48\)).

Results show that male pre-service teachers have more self-confidence than female pre-service teachers in terms of using technology in mathematics education. However, it indicates that female students got higher scores than males in terms of the results of the perception survey towards technology. Besides, freshman students have lower self-confidence than the sophomore and senior students in terms of using technology, whereas junior students have less confidence than senior students.

**Discussion and conclusion**

The purpose of this study was to investigate TPACK, TPACK related self-confidence, and the perception of pre-service teachers based on the use of instructional technology in terms of gender and grade level. Results showed that male participants had higher scores on the surveys of TPACK and TPACK related self-confidence than female participants. In contrast, female participants had higher scores on the perception survey towards technology use than their male counterparts. This result coincides with the results of other studies stating that there is a significant relationship between gender and technology perception (Gilley, 2002; Siyambas, 2015; Tnmaz, 2004). The study of Jamieson-Proctor, Finger, and Albion (2010) emphasized that self-perception of the pre-service teachers based on their competence with instructional technologies proposed that they did not have a high level overall. Within the current study, female participants had positive attitudes towards technology use in mathematics education. Although the comparison of the scores of pre-service teachers shows that there was no significant difference between the groups of male and female participants on each survey, the total scores were in favor of male participants. This result coincides with the results of the previous studies conducted on 49% female and 51% male participants (Durndell & Haag, 2002) and 79% female and 21% male participants (Jamieson-Proctor, Finger, & Albion, 2010). In those studies, it was found that male pre-service teachers had higher self confidence to use technology in education. Especially in the study of Jamieson-Proctor, Finger, and Albion (2010), it was found that male teachers were very confident to use instructional technologies while female teachers had no or little confidence. Therefore, male and female teachers differ in terms of their confidence to use the instructional technologies (Jamieson-Proctor, Finger, & Albion, 2010). The results may suggest that male pre-service teachers are more confident regarding the use of educational software for teaching of mathematics. This might be caused by the widespread use of technology by male pre-service teachers than
female counterparts. That is, this result might be related with the readiness of pre-service teachers to use instructional technologies rather than the teacher education program, and male pre-service teachers may feel more competent to use technological software than their female colleagues. In order to increase the self-confidence of both male and female pre-service mathematics teachers, different types of instructional activities might be presented to pre-service teachers by the use of instructional technological software in methods for teaching courses or other elective courses related to technology use in mathematics education. Moreover, pre-service teachers might be encouraged to create instructional activities including technological software on their own. Therefore, they may become more confident since they participate in the process of producing such activities. As a result, pre-service teachers might gain experience of the use of instructional technologies in real classroom environment.

The results for the analyses with respect to the grade levels were different from the study of Jamieson-Proctor, Finger, and Albion (2010). Although they found no difference based on the age of the participants, freshman pre-service teachers had significantly lower scores than the senior pre-service teachers within the current study. Therefore, it might be inferred that undergraduate courses such as Methods of Teaching Mathematics given at fifth and sixth semesters and Computer-based Mathematics Education given at seventh semester in teacher education program get pre-service teachers improve their self-confidence to use technology in mathematics education. On the contrary, there was not a consistent change from freshman to senior pre-service teachers in the current study since there was a break on the results of junior pre-service teachers. By the participation of more pre-service teachers, the relationship among grade levels might be observed more precisely for the same variables. In addition, since already existing groups were used in the current study, the results might be biased. Therefore, further studies might be conducted by increasing the number of participants and by the use of randomization.

As Jamieson-Proctor, Finger, and Albion (2010) stated, teacher education programs should ensure that all pre-service teachers have the necessary knowledge bases such as technological knowledge and TPACK to integrate instructional technologies into education. In this respect, before the completion of teacher education program, the improvement of knowledge and self-confidence of pre-service teachers in order to use instructional technologies is a favorable outcome. Furthermore, the perception of pre-service teachers is a crucial element to integrate technology into instruction (Parr, 1999). Therefore, positive attitude of pre-service teachers towards instructional technologies was one of the positive outcomes of this study. In addition, a relationship was observed between the use of technology and self-confidence towards the use of technology as previous researchers found (Marakas, Yi, & Johnson, 1998; Wilfong, 2006). Therefore, the courses provided for pre-service teachers and in-service teachers might get them confident to use instructional technologies in mathematics education. As it is stated above, the TPACK, TPACK related self-confidence, and perception towards the use of instructional technologies of pre-service mathematics teachers might be improved by getting them participate in the creation and using procedure of learning activities which includes instructional technologies. If students engage in such activities adequately, they might appreciate the value of the use of instructional technologies in mathematics education. Therefore, their knowledge, self-confidence, and perception might be changed in this way.

Consequently, the study was limited to 427 pre-service middle school mathematics teachers from three different universities, who were predominantly female. Also, the results of the study were limited with the data obtained from the questions of three surveys. Hence, the generalizability of the conclusions of this study to larger student populations or other contexts might be limited. In addition, the results of the study were limited with quantitative analyses of data. Therefore, qualitative analyses might also be used in order to investigate the context more deeply. Lastly, empirical studies might be conducted related to the same research problem in order to control several factors and explain the results more effectively.

References


130


Social and Collaborative Interactions for Educational Content Enrichment in ULEs

Rafael D. Araújo1*, Taffarel Brant-Ribeiro1,2, Igor E. S. Mendonça1, Miller M. Mendes1, Fabiano A. Dorça1 and Renan G. Cattelan1

1Faculty of Computing, Federal University of Uberlândia, Uberlândia, MG, Brazil // 2Federal Institute of Education, Science and Technology of Southern Minas Gerais, Passos, MG, Brazil // rafael.araujo@ufu.br // brant.ribeiro@ifsuldeminas.edu.br // igoremendonca@gmail.com // m3iller@gmail.com // fabianodor@ufu.br // renan@ufu.br

*Corresponding author

(Submitted February 10, 2016; Revised August 9, 2016; Accepted October 13, 2016)

ABSTRACT

This article presents a social and collaborative model for content enrichment in Ubiquitous Learning Environments. Designed as a loosely coupled software architecture, the proposed model was implemented and integrated into the Classroom eXperience, a multimedia capture platform for educational environments. After automatically recording a lecture in instrumented classrooms, students and instructors can enrich its content with comments and rating features. The platform usage was monitored for three school semesters to analyze the receptivity and the impact of the proposed features over 121 undergraduate students. As a result, we observed that both system’s access rate and students’ performance increased, suggesting that interactive features leverage collaborative learning interactions and promote the teaching/learning process.

Keywords

Educational content enrichment, Social and collaborative learning, Ubiquitous learning environments

Introduction

Recent surveys have revealed that the number of computing devices, especially mobile ones such as smartphones and tablets, has grown to a great extent in last years (Gartner, 2014), and their usage for performing everyday tasks is now a reality (Kostakos & Ferreira, 2015). Technological advances coupled with infrastructure provided by the Internet provide a scenario in which information can be accessed anytime and anywhere. This new paradigm of interaction between people and computers is known as Ubiquitous Computing (UbiComp) (Weiser, 1991), and illustrates the omnipresence that technology reached in different environments, including schools and classrooms.

In the educational scenario, UbiComp systems are able to assist instructors and students during the teaching/learning process by automating pedagogical tasks, thus creating the so called Ubiquitous Learning Environments (ULEs) (Settle, Dettori, & Davidson, 2011). Instrumented classrooms are the most common approach for this purpose. Devices, such as electronic whiteboards, cameras, and microphones, can produce media artifacts that are able to recreate experiences that took place in the classroom. In this way, students can focus their attention on the learning experience itself, certain that details are being properly recorded and will be available for later access (Brant-Ribeiro, Cattelan, & Biase, 2015).

ULEs are quite related to Capture and Access (C&A) applications (Truong & Hayes, 2009), a recurring research theme in UbiComp. Supported by C&A, ULEs promote automatic authoring of multimedia content. In this context, social and collaborative interactions constitute significant functional improvements for classic ULEs (Banday, 2012; Cela, Sicilia, & Sánchez, 2015). With these systems, content presented by instructors can be extended and enriched, which encourages information sharing, experiences exchange and discussion among students who start performing active and influential roles in cognitive processes of the members of virtual communities where they are inserted (Banday, 2012; Shukor, Tasir, Van der Meijden, & Harun, 2014).

Therefore, we developed a model for extension and classification of multimedia content based on social and collaborative activities carried out by users in ULEs. As a proof of concept, we proposed and implemented a software architecture that supports assumptions of the proposed model, resulting in an application that provides features for commenting and rating multimedia artifacts. We conducted a case study with integration and validation of the collaborative application into the Classroom eXperience (CX) (Araújo, Brant-Ribeiro, Cattelan, de Amo, & Ferreira, 2013) – a UbiComp platform built to automatically capture and access educational activities in instrumented classrooms.
In addition, we also performed a comparative analysis among our proposal and related studies that include solutions for extension of multimedia content, categorization of digital artifacts and collaborative annotations. Our study, however, tries to meet all these demands through a generic model developed to support the enrichment of media components based on collaborative authoring and rating, and allows analysis of resulting applications at distinct levels of granularity.

**Multimedia content enrichment model**

The continuous knowledge progress coming from different educational activities is the basis of the spiral content production model (Pimentel, Ishiguro, Kerimbaev, Abowd, & Guzdial, 2001). In this model, the spiral represents all generated information in each context (program, course, class) and each pedagogical activity contributes toward the new development cycle. This way, the spiral’s area is increased by instructors and students’ interactions and the produced knowledge is leveraged.

The proposed multimedia content enrichment model aims at supporting the task of associating and categorizing media artifacts in Web environments in a simple and concise way. Its structure was designed to expand and to classify the associated content based on information provided by users. Also, the model fosters hierarchical digital artifacts construction, which encourages its use by personalization and recommendation systems.

This approach is based on Composite Hypermedia (Ismail, 2009; Khan & Tao, 2001), a model for building multimedia documents in which different media can be combined to create new hypermedia artifacts with their own characteristics. The proposed model contains three types of media, named “Artifact”, “Comment”, and “Rating.” An Artifact is the main object and it refers to the digital content that will be extended with information from other components. A Comment is textual information that can be associated to either an Artifact or other Comments. Lastly, a Rating represents the users’ acceptance regarding those components.

In addition to the components, association rules among them are also defined: (i) an Artifact may be associated only with other Artifacts; (ii) a Comment may be associated with either Artifacts or Comments; and, (iii) a Rating may also be associated with Artifacts and Comments. Figure 1 depicts an abstraction of the collaborative model, in which rectangles represent components and arrows indicate their relations, i.e., which components can be linked to others, according to established rules. As can be seen, only Artifacts have a recursive association and they are also the only components that are able to communicate directly with the collaborative platform.

![Figure 1. Multimedia content enrichment and rating components, adapted from Brant-Ribeiro et al. (2014)](image)

The model takes into account two sources of information, called entities, which can be the Host Platform or the User. The Host Platform produces artifacts that will be enriched by means of comments and ratings made by users who share common interests, which characterizes the social nature of this model. Also, it is called collaborative because produced information extends the already existing content.

Therefore, relationships among components of this model can be used as a source of information to entities that implement content recommendation and personalization algorithms. In addition to their natural relationship, comments and ratings can be employed as parameters for helping decision making of recommendation algorithms.

**Case study with the Classroom eXperience platform**

Classroom eXperience (CX) (Araújo et al., 2013; Brant-Ribeiro et al., 2014) is a C&A platform deployed at the Faculty of Computing at Federal University of Uberlândia (FACOM/UFU) that was developed as an educational support tool to record classes in instrumented settings, store captured multimedia content, and make it available...
to students for further revision. CX automatically generates hypermedia documents in different presentation formats by synchronizing media streams coming from ubiquitous devices, such as electronic whiteboards, cameras, and projectors.

Its architecture follows a well-structured sequence of activities and implements the model proposed by Truong and Hayes (2009), which is composed by four phases, namely pre-production, live recording, post-production, and access. Besides, CX expands its scope to explore an additional phase, called extension (Pimentel et al., 2001), in which the content is enriched continuously by users’ interactions, as shown in Figure 2.

![Figure 2. C&A process, adapted from Brant-Ribeiro et al. (2014)](image)

A prototype of the proposed collaborative model was built as a loosely coupled module on top of the CX’s Web front-end, providing extensibility and reusability properties, which has not changed the CX’s operation.

**Design and implementation**

The prototype’s architecture follows the principles of Domain-Driven Design approach, in which domain-specific features are built and independent of each other (Santos, Beder, & Penteado, 2015). Thereby, Web services were built to enable the communication between the two modules: one for performing service requests and display graphical components, and another for fetching and storing system information.

Lectures’ slides rating is one of the implemented features which uses the star rating approach that is commonly employed in Web platforms. Also, the average rating of each slide is presented by another visual component. Once users measure slides importance by rating them, this information can be employed both as a customization parameter as well as input for recommendation algorithms.

Students can also create comments for lectures’ slides and the course itself. Comments are eligible for replication and rating, which encourage debate and can be used to measure comments relevance. However, comments’ rating uses a different approach from slides rating. For comments, the “thumbs-up/thumbs-down” approach that is commonly used in social networks was adopted. Figure 3 depicts the content presentation page in CX, in which region (a) presents the rating component, region (b) depicts the slides comment component, and region (c) highlights the slides rating average component. Also, the video component can be noticed at the figure’s top right corner and the slides navigation mechanism at the bottom.

To illustrate an instance of the slides rating service, Figure 4 depicts the steps required to perform a slide rating registration and fetching as well as the relationships among the architecture components. When a user request for a slide (first step), an Ajax request triggers the rating registering Web service (second step). Upon the request receiving, the artifact rating is stored by a Web service (third and fourth steps) and, then, sends a JSON object back with the artifact rating information (fifth, sixth, and seventh steps). Finally, the HTTP response is handled...
Information gathered by the collaborative module enriches the existing digital content, fostering collective learning, and producing a social network made up of students and instructors. Since some user information is also kept by the module, it is capable of generating input for expert systems based on users’ interactions.

**Evaluation method**

An experiment employing an adaptation of the Technology Acceptance Model (TAM) (Davis, 1986) was conducted in order to analyze the impact of social and collaborative features in CX, and also to verify their acceptance by users evidencing which factors most affected the system usage intention. TAM aims at understanding the users’ acceptance process of computer-supported technologies and it may be tailored to fit each application context. Several works that make changes and extensions of that model are found in the literature (Peris, Blinn, Nüttgens, Lindermann, & von Kortzfleisch, 2013; Rauniar, Rawski, Yang, & Johnson, 2014).

Two original TAM dimensions related to motivational issues were employed as a way of predicting the actions of individuals, namely Perceived Usefulness (PU) and Perceived Ease of Use (PEU). PU concerns the degree that
users believe the usage of the new technology can improve their performance or usage experience. PEU can be understood as the level of effort that users believe to be necessary for using the new technology.

The following hypotheses were drawn up based on those dimensions, also considering the context of use of CX:

**Hypothesis 1**: The PU of both the CX platform and collaborative features directly influence users intended usage.

**Hypothesis 2**: The PEU of both the CX platform and collaborative features directly influence users intended usage.

In order to increase the analysis reliability and explore specific factors of the system, two dimensions have been added to the analysis criteria: Perceived Attractiveness (PA) and Academic Performance (AP). PA is often found in studies that apply TAM in Web-based platforms. Such criterion might be understood as the attractiveness level of applications’ components and functionalities. Its analysis intends to identify the extent to which this relationship affects the interest in the system’s usage. As a result, PA criterion was included among the evaluated dimensions and the following hypothesis was analyzed:

**Hypothesis 3**: The PA of both the CX platform and collaborative features directly influence users intended usage.

Finally, the adoption of information technology in the educational context aims at promoting teaching/learning processes by providing positive results for students and instructors. In this way, AP dimension has been added to the criteria set to evaluate students’ academic performance, and the following hypothesis was drawn up:

**Hypothesis 4**: The use of both the CX platform and collaborative features affect the AP of users who had used the educational system.

For investigating those dimensions, evaluation questionnaires were administered and students’ access logs were analyzed. Students from five undergraduate classes – bachelor’s degree in Computer Science and in Information Systems at FACOM/UFU – were part of the experiment during two academic terms. The questionnaire was designed to obtain users’ feedback about the collaborative features based on the proposed TAM adaptation. In order to ensure that thoughtless answers would not influence the research outcome and to neutralize inconsistent answers, a method for reversing and negating statements was employed during the questionnaire design (Huang, Curran, Keeney, Poposki, & DeShon, 2011; Weijters & Baumgartner, 2012).

Thus, two assertions for each dimension were prepared at first. In addition, two more assertions for each dimension were prepared with the opposite idea to the first ones (negating technique), leading to a 16-statements questionnaire containing different points of view about CX’s usability and collaborative features, and also their impact in the analyzed context. For each statement, it was provided a 5-points Likert scale representing the user agreement degree about that item, ranging from “Strongly disagree” to “Strongly agree”. Employed statements are presented below:

- I liked to rate slides with stars, expressing their relevance.
- I accessed only particular lectures when I logged in CX.
- I think user comments provided additional information to the slides content.
- Slides rating information, like the overall mean of stars, was more useful to study than the user comments.
- In my opinion, comments and ratings (stars) helped me to understand the presented content.
- I found it tough to use the interface and confusing how to create comments.
- Stars rating was intuitive and simple to use.
- In the user comments, I did not find relevant information that helped me to understand the slides.
- I think that rating slides with stars is not relevant.
- I accessed different lectures each time I logged in CX to study/review.
- For me, comments were not relevant and did not offer extra information about the slides.
- In my opinion, comments were more relevant to the lectures than slides ratings (stars).
- I think the slides comments and ratings (stars) features did not help in my learning.
- I liked to interact with CX interface and it was easy to create and read the user comments.
- Star rating got a bit confusing and it was tough to use.
- User comments helped me to understand the course content.
The link between questionnaire statements and TAM dimensions was as follows: statements 1 and 10 were about PA positive aspects while statements 2 and 9 had opposing ideas; statements 3 and 4 and their negated ones (11 and 12) described PU; statements 5 and 16 and their opposing ones (8 and 13) were about AP; and, finally, statements 7 and 14 and their opposing pairs (6 and 15) were related to PEU.

Additionally, a field for comments and suggestions was also provided, assuming that those responses would help to measure the satisfaction level of users’ experience, indicating points of failure that could be considered for forthcoming platform evolution. Besides administered questionnaires, access logs of students who attended the two courses that used CX as a support tool were also analyzed.

Results and discussion

During the school semesters when the social and collaborative features were integrally available in the CX platform, 77 students utilized them. Of these, 25 answered our proposed questionnaires. For each pair of affirmatives with the same idea, we expected students to respond with opposing impressions, since these statements had different views about the same concept. Thus, validation of statements was carried out by inverting the item that had the negative idea of each pair and performing the calculation of grouped results, in order to confirm or cancel responses for each idea. Results achieved with the use of this legitimization technique can be observed in Figure 5, which presents the agreement and disagreement levels (in percent) of students in relation to the affirmatives.

![Figure 5. Student agreement levels for the questionnaires’ affirmatives](image)

In general, students agreed with the positive ideas presented in questionnaires. Affirmatives 1 and 10, which sought to measure student PA level in relation to CX, reached an elevated level of concordance. This demonstrates that users considered attractive the functions present in CX and, thus, were encouraged to navigate the system and access different recorded classes. Most of students who answered the questionnaire also reported that CX’s user experience was satisfactory, since it presented an intelligible interface.

We also perceived that users acknowledged the social and collaborative features as sources of additional information which assisted them to understand the content presented in classroom. These facts can be concluded through the analysis of users responses to affirmatives 7 and 14, which comprised users PEU in relation to the platform, and 5 and 16, which ascertained if students considered the collaborative features as ways of obtaining extra content beyond the classroom. Concerning the PU dimension, analysis of affirmatives 3 and 4 revealed that students recognized the social and collaborative components as useful features in the platform.

Responses obtained through the reserved field for criticism and suggestions were also analyzed. There were some comments about CX unavailability in some periods and the possibility of downloading recorded content for offline study. Platform unavailability was due to casual instabilities of the university’s network where this research was conducted, a factor beyond the reach of the platform’s maintainers. About providing downloadable material for users, since CX explores the dynamics of UbiComp over existing models based on static documents, we believe this is not an issue, but an expected part of students’ adaptation process. Furthermore, students pointed out that slides classification and written comments had a great value for both the interaction among peers and the recorded content extension with didactic materials for later study.
In addition to questionnaires, we also analyzed logs of usage and collaborative activities. We observed the access frequency of five classes enrolled in two courses (A and B), which used CX as a support tool. Classes’ first semester represented the control group, while classes’ second semester used the social module. Data was plotted to visualize the evolution of weekly accesses in the system as well as in the captured lectures. Figure 6 depicts the amount of accesses performed by students that attended course A (in both classes). Classes that attended course B presented similar access levels.

Access peaks observed in Figure 6 illustrate weeks that students had tests. In these periods, CX received higher amounts of hits than in other weeks, a result that allows us to comprehend that students used the system to study for exams. In addition, Class 2 (which employed CX integrated with the social module) had hits peaks that exceeded Class 1 in over a hundred visits to lectures during tests periods. Therefore, we believe that students who utilized CX with the social module felt more encouraged to interact with the captured content by performing various slides ratings and comments during every visit to the platform.

We also analyzed student grades to understand whether the inclusion of new technologies brought gains to students’ performance. To accomplish this analysis, statistical tests were conducted to strengthen if there were significant differences among students’ grades. In order to employ appropriate sample sizes to correctly analyze student performances, we utilized a minimum of 20 observations, as established in Brant-Ribeiro and Cattelan (2015). Also, to avoid result biases, we disregarded grades of students that, at the end of the semester, presented dissonant performances if compared to other grades in the same academic periods. We understand that these students had higher chances of careless behavior at the end of the semester, since they had already reached the minimum requirements to be approved in the course.

Samples were initially submitted to the Shapiro-Wilk’s test (SW) in order to ascertain if the statistical residuals of the observed variable followed a normal distribution. Subsequently, Levene’s test was applied to samples from different classes who attended the same courses to check the homogeneity of variances between them. Then, the Student’s t test was utilized to verify if there were significant differences among means of students grades who attended the same courses in different classes. Table 1 presents the results of SW, Levene and Student’s t tests, as well as the respective performances of analyzed groups.

Table 1. Classes’ performances with and without the social features

<table>
<thead>
<tr>
<th>Course</th>
<th>Class</th>
<th>n</th>
<th>SCM</th>
<th>$\bar{X} \pm s$</th>
<th>$W(P)$</th>
<th>$F(P)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>21</td>
<td>Absent</td>
<td>78.85 ± 12.89 a</td>
<td>0.961 (0.533)</td>
<td>0.284 (0.597)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>35</td>
<td>Present</td>
<td>80.19 ± 11.62 a</td>
<td>0.960 (0.231)</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>23</td>
<td>Absent</td>
<td>62.16 ± 21.27 a</td>
<td>0.953 (0.340)</td>
<td>4.702 (0.036)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>20</td>
<td>Present</td>
<td>74.84 ± 14.62 b</td>
<td>0.927 (0.138)</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>23</td>
<td>Absent</td>
<td>62.16 ± 21.27 a</td>
<td>0.953 (0.340)</td>
<td>1.665 (0.204)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>22</td>
<td>Present</td>
<td>74.80 ± 15.02 b</td>
<td>0.921 (0.080)</td>
<td></td>
</tr>
</tbody>
</table>

Note. 1Means followed by distinct letters in each course differ from each other through the Student’s t test for independent samples with 0.05 of significance; SCM: Social and Collaborative Module; $\bar{X} \pm s$: Mean and standard deviation; $W$: Statistic of Shapiro-Wilk’s test; $F$: Statistic of Levene’s test; $(P)$: Probabilities above 0.05 indicate statistical residuals with normal distribution and homogeneity of variances for Shapiro-Wilk’s and Levene’s tests, respectively.
Every observed sample exhibited residuals’ normality. Only between classes 3 and 4 of course B it was not possible to achieve homogeneity of variances through Levene’s test, but the degrees of freedom were adjusted for this case and, therefore, it was possible to employ Student’s t test to compare these samples. Despite obtaining higher grades in every class that employed the social module, only students who attended course B presented a real performance increase in their grades. In this course, besides the means increase, there were also standard deviations diminution – what allows us to understand that students of the analyzed classes started studying from common information sources, generating a higher flattening of grades.

We also analyzed grades of students who had partial contact with CX’s social module. The one-sample Student’s t test was employed to compare grades of students in the same academic period, since the features had been introduced in CX between the 13th and 14th weeks of the semester. Thus, students had contact with the social features to study for the final tests that took place between the 16th and 17th weeks of the semester. For the one-sample Student’s t test, we subtracted the student performance obtained in the period when there was no social module from the grades the same students achieved when the social features were available. The SW test was applied to these results and we observed that all samples followed a normal distribution. Finally, the samples were submitted to the Student’s t test. Table 2 presents the results of analysis made in both classes during the same semester.

<table>
<thead>
<tr>
<th>Course</th>
<th>n</th>
<th>Bimonth</th>
<th>SCM</th>
<th>Mean ± SD</th>
<th>Mean difference ± SD</th>
<th>W (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21</td>
<td>1st</td>
<td>Absent</td>
<td>39.42 ± 6.45</td>
<td>0.38 ± 8.28</td>
<td>0.979 (0.908)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>Present</td>
<td>39.81 ± 7.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>23</td>
<td>1st</td>
<td>Absent</td>
<td>30.86 ± 10.56</td>
<td>7.76 ± 12.41</td>
<td>0.984 (0.958)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>Present</td>
<td>38.62 ± 8.73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Means followed by distinct letters in each course differ from each other through the one-sample Student’s t test with 0.05 of significance; SCM: Social Collaborative Module; Mean and standard deviation; ∆ ± s: Mean difference and standard deviation of difference; W: Statistic of Shapiro-Wilk’s test; (P): Probabilities above 0.05 indicate statistical residuals with normal distribution for Shapiro-Wilk’s test.

Analysis of Table 2 allows the understanding that there was again a difference only among students who attended course B. With regards to course A, even also presenting a visible mean increase, there was no significant difference among student grades. Therefore, analysis of classes for both full and partial utilization of the features revealed that only students who attended course B presented an increase in their performances. We believe this happened because course B had a mathematical focus, in which CX resources were extensively employed. Since the aforementioned course dealt with complex subjects, in which most of the problem resolutions demanded higher commitment by enrolled students, this suggests that courses characteristics greatly influence on how receptive is the technological support offered by CX to users.

In short, from the analysis of user questionnaires responses and observation of CX access logs, we found out that PU, PEU and PA of students were positive and consistent. Issues such as usability, attractiveness, and enjoyment were evidenced by students who used the platform. Thus, it was possible to comprehend that CX was perceived as a useful supporting tool to the educational context and that it has attractive features to users, such as an intuitive interface. In this way, the hypotheses 1, 2 and 3 were altogether supported by the CX platform.

Also, analysis of students’ performances provided the understanding of a significant improvement in their grades among students who attended course B. In this context, hypothesis 4 was supported with reservations, since the expected improvement for this topic was partially supported in this research application context. However, the analysis about the influence of courses characteristics upon students along with the support of educational technology presented positive results and aspects in this research. This factor evidences that hypotheses which deal with student grades must include issues and questions that go beyond the single usage of new technologies. This way, the validation of this study’s guidelines employing the TAM model was efficient for the analysis of the impact and acceptance of social and collaborative features in CX. Adjustments made in the TAM model to fit our context were appropriate and generated satisfactory results.

Related work

Several studies concerning to UbiComp in educational settings can be observed in the literature. The use of interactive technologies by academics opens doors to a new era of teaching and learning, breaking down barriers that hindered the full achievement of knowledge. Typical researches in this area deal with recommendation
Collaborative learning is a methodology focused on the process of getting knowledge through social interactions, either in real or virtual environments. A common goal of educational approaches is to foster communication and teamwork skills among students (Mukherjee, Pal, Choudhury, & Nandi, 2014). Encouraging group problem solving is one way of achieving those goals.

Claros and Cobos (2013) tackle this theme by presenting the Social Media Learning, an educational platform that supports interactive construction of multimedia content. Content rating, management of educational content and multimedia authoring are features of the platform. In addition, they get social information from Facebook, and use Youtube services for videos delivering. However, this platform does not classify Learning Objects (LOs) in hierarchical levels and they do not provide content personalization mechanisms, as in CX.

A collaborative learning environment called Collaborative Science Inquiry (CSI) was presented by Sun, Looi, and Xie (2014). It consists of two functional modules, Teacher Module and Student Module, that allow teachers to design instructions and questions, attach simulations, manage groups, and review learning artifacts, and students to collaborate by means of inquiry activities in a shared workspace. Although it has a collaboration feature, the content is neither classified hierarchically nor personalized.

Mukherjee et al. (2014) explore the concept of Mobile Learning by proposing a Mobile Ad Hoc Network (MANET) infrastructure for setting up a collaborative educational environment. In that environment, mobile devices allow users to communicate directly with each other. Despite of promoting social interactions and discussions, approaches that consider only mobile devices may have technical constraints, such as coarse-grained content viewing due to the small size of the devices’ screen. Furthermore, the way of obtaining and providing data is different from the proposed approach since it explores the Web 2.0 features.

Chen, Hwang, and Wang (2012) created MyNote, a system that promotes collaborative construction of knowledge by means of authoring and sharing of notes for multimedia artifacts in Web 2.0 environments. It allows to create comments on notes and classify them in order to foster debates and exchange of information. MyNote’s architecture is similar to the one presented in this article, since it has components embedded in the host platform that are responsible for calling Web services. Nevertheless, they store the content in a different way by carrying out a copy of the entire HTML code (raw page and notes) along with its URL and the notes’ timestamps, while in CX the notes are linked to the multimedia artifact itself, regardless the way they are presented to the users.

Fan et al. (2010) developed a Web platform that provides tools for creating notes on images. It allows users to create text or multimedia annotations in specific areas of the images, which may facilitate further discussions and analysis of figures and graphs. Azouaou, Mokeddem, Berkani, Ouadah, and Mostefai (2013) created WebAnnot, a Web tool for creating, categorizing and sharing online notes, which uses an ontology-based model that has standard properties such as attributes that allow the search and recommendation of semi-automatic or manually created annotations. Yet, WebAnnot does not support neither classification of annotated multimedia artifacts nor their hierarchy.

In the work of Morgado, Penalvo, and Hidalgo (2012), they presented a methodology for categorization and classification of LOs considering students’ skills and abilities. This way, it is possible to present LOs suited to students with different skills. However, the inference process for LO categorization is done manually by instructors and, thus, collaborative aspects are not exploited.

Foll, Pontow, Linner, and Radosch (2006) developed a conceptual framework for building applications to classify and manage multimedia content produced by social and collaborative activities. It includes features such as content management, collaboration, personalization and recommendation based on contextual and social activities. They implemented a Web blog, called Online Community Life (OCL), which enables the distribution of multimedia content in communities managed by ubiquitous platforms. However, that approach does not provide mechanisms that enable users to create and manage their social and collaborative information within those ubiquitous platforms that are integrated with the OCL.

In order to better illustrate related work, Table 3 summarizes the information presented in this section. For comparison purpose, we used the main features supported by the proposed social and collaborative model. In this way, the following criteria for evaluation were highlighted: (a) support for enrichment of multimedia artifacts;
(b) support for classification of multimedia artifacts; (c) support for collaboration among users (sharing and discussion); (d) hierarchy of artifacts, and (e) support for content personalization based on collaborative activities. For each criterion, three levels of compliance were determined: an unfilled circle (○) for works that do not support the evaluated criteria; a half filled circle (●) for works that partially support the evaluated criteria and fully filled circle (●) for works that support the evaluated criteria.

Table 3. Comparative table among the proposed and related work

<table>
<thead>
<tr>
<th>Work</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Claros &amp; Cobos, 2013)</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(Sun et al., 2014)</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(Mukherjee et al., 2014)</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(Chen et al., 2012)</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>(Fan et al., 2010)</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>(Azouaou et al., 2013)</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>(Morgado et al., 2012)</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>(Foll et al., 2006)</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>This study</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

Some approaches have their focus on multimedia content enrichment, while others go deeper in the classification of artifacts. In most studies, collaborative authoring is also taken into account as way to encourage the collective constructivism and improve results against user satisfaction. However, our approach differs those by presenting a generic model for annotations and their types and relationships in addition to considering content enrichment and classification (ranking) requirements, supporting content hierarchy at different levels.

Conclusion and future work

The approach discussed in this paper sought to explore social and collaborative features within real ULEs. To do so, we proposed a software architecture model capable of supporting collaborative interactions for content extension and enrichment via annotations and classification (ranking). Such architecture also provides support for content recommendation and personalization. Based on the proposed model, a collaborative system was implemented to create annotations and classify multimedia content.

As a case study, the proposed approach was integrated into the Classroom eXperience (CX) platform as an add-on module. For the sake of validation, the Technology Acceptance Model (TAM) adapted to the educational context was carried out. Our experiment comprised the administration of questionnaires to obtain users’ point of view, analysis of access logs, and study of students’ grades. Results demonstrated that features offered by the collaborative module were well accepted by users and brought some benefits, such as higher number of access to the system and higher final grades. We found that the social features have encouraged CX’s usage and fostered collaborative learning among students, which is an important resource for enriching the content beyond what was discussed in class. In addition, customized TAM revealed itself appropriate for the validation of this proposal.

Although including social and collaborative features in CX has proved to be valid for the students investigated in this study, we cannot generalize the results because they may be biased by the domain of the studied courses (Computer Science and Information Systems). Students of those courses might have a greater willingness to accept new technologies than students of adverse areas. This assumption leads to the need for a larger and more diversified sample, considering an interdisciplinary scope, so that it is possible to obtain results in diverse technological axes and achieve greater conviction about the importance of developing and studying social and collaborative features in educational systems.

Also as future work, gamification elements have been investigated in the context of the CX platform as a complementary approach to increase students’ motivation. Moreover, social information obtained through the proposed model will also be used to adjust the students’ model, which may include cognitive aspects such as their learning styles, and as input for recommendation and personalization algorithms.

Finally, we believe that the exploitation of UbiComp premises in the educational field remains as a potential alternative to refine the teaching/learning processes since it aims to assist people through applied technologies in a non-intrusive way. Such approaches match the construction of environments in which the real and virtual worlds are blended in a way to support the use of information technology.
Acknowledgements

The authors would like to thank CAPES, CNPq, FAPEMIG, PROPP/UFU and PET/MEC/SESu.

References


Students’ Metacognition and Cognitive Style and Their Effect on Cognitive Load and Learning Achievement

Omar López-Vargas1, Jaime Ibáñez-Ibáñez1 and Oswaldo Racines-Prada2

1School of Technology, Universidad Pedagógica Nacional, Bogotá, Colombia // 2Secretary of Education of Bogotá, Bogotá, Colombia // olopezv@pedagogica.edu.co // jibanez@pedagogica.edu.co // oswaldoracines@gmail.com

Corresponding author

(Submitted February 20, 2016; Revised May 11, 2016; Accepted June 26, 2016)

ABSTRACT

The present research’s objective is to examine the effects of metacognitive scaffolding and cognitive style in the Field Dependence - Independence (FDI) dimension on cognitive load (CL) and learning achievement (LA) in high school students, when they interact with a hypermedia environment on philosophy (logic). Fifty-four students belonging to two eleventh grade courses from a public school in Bogotá - Colombia participated in the study. One of the student courses interacted with a hypermedia environment that contained, within its structure, the metacognitive scaffolding. The other course interacted with the hypermedia environment that did not have the scaffolding. Students were given the Embedded Figures Test (EFT) to classify them into field dependent, intermediate, and independent subjects. A Repeated Measures Analysis was conducted with two intra-subject variables: (1) CL and (2) LA. Findings indicate that significant differences exist between intrinsic and extraneous cognitive load because of the effect of the metacognitive scaffolding. Students that interacted with the metacognitive scaffolding exhibited significantly greater achievements than those that did not use it. The field independent students also exhibited significant differences in CL with respect to their field independent and intermediate classmates.

Keywords
Metacognitive scaffolding, Cognitive load, Cognitive style, Learning achievement, Hypermedia environment

Introduction

In recent decades, different computer-based learning environments (CBLE) have been used in an educational context to provide support for the teaching-learning process in different levels of schooling. The use of these environments in the classroom has generated high expectations among the academic community since it is believed that when students interact with these scenarios, they can take on a more active role in their own learning process and thus achieve more successful and motivating learning experiences (Clark & Meyer, 2008; Mayer, 2005; McNamara & Shapiro, 2005; Shapiro, 2008). However, some studies indicate that little empirical evidence exists to support these expectations since in some cases, students do not accomplish the desired learning nor do they all equally benefit from these environments (Alomyan, 2004; Beserra, Nussbaum, Oteo, & Martin, 2014; Calandra & Barron, 2005; López-Vargas, Hederich-Martínez, & Camargo-Uribe, 2012).

In this research field, some studies explain that LA obtained by students when interacting with computational environments may be directly related to student’s cognitive style and CL. Regarding cognitive style, for example, in the Field Dependence - Independence –FDI- dimension, most of the studies show that field independent novices, when interacting with hypermedia environments, organize and process information more efficiently and obtain greater LAs in comparison to their field dependent classmates (Alomyan, 2004; Altun & Çakan, 2006; Chen & Macredie, 2002; Handal & Herrington, 2004; López-Vargas & Valencia-Vallejo, 2012).

With respect to CL, some research show that the characteristics in the design of computational environments can favor or limit students’ learning process. Thus, the mental effort employed by a subject when developing a learning task may be negatively affected if the organization of the information presented overloads the limited resources of the working memory. This situation efficiently affects knowledge building (Artino, 2008; Clark & Mayer, 2008; Mayer, 2005; Sweller, Ayres, & Kalyuga, 2011; Sweller, van Merrienboer, & Paas, 1998).

On the other hand, studies show that the use of scaffolding favors subjects’ performance when they undertake learning tasks in an autonomous manner in computational environments (Greene, Moos, Azevedo, & Winters, 2008; Delen, Liew, & Willson, 2014; Kim & Hannafin, 2011; Lehmann, Hähnlein, & Ifenthaler, 2014; Zhang, 2013). In this research area, the use of metacognitive scaffolding in computational environments is an aid for the student when managing and regulating cognitive processes during the learning process. Thus, the subject plans activities, monitors and controls the progress of proposed goals, and evaluates the obtained results (Molenaar, Boxtel, & Sleeegers, 2010; Quintana, Zhang, & Krajcik, 2005; Zhang, 2013; Zhang & Quintana, 2012).
Literature review

Field Dependence - Independence (FDI)

In an educational context, the most studied cognitive style is the Field Dependence - Independence (FDI) dimension proposed and developed by Witkin and his colleagues (Witkin & Goodenough, 1981). In an information technologies context, research on cognitive style in the FDI dimension systematically show that students referred to as Field Independent (FI) obtain better LAs than their Field Dependent (FD) classmates when interacting in hypermedia environments. Studies evidence that FD students prefer their study material to be organized sequentially (linear) since they are easily disoriented and they do not know where to begin, nor in what direction to continue; situation that makes it harder for them to effectively structure and restructure the information. Additionally, they prefer the browsing process in the computational scenario to be in groups and guided by external agents, and that the control over the learning process be exercised by the own computational environment (Alomyan, 2004; Chen & Macredie, 2002; Handal & Herrington, 2004).

In contrast, FI students prefer autonomy to browse throughout the whole structure of the computational environment and effectively handle hypermedia environments. They can establish browsing routes in a structured fashion. Similarly, while browsing they are not easily distracted with irrelevant information and they can effectively use most of the computational environment’s resources. On the other hand, they like to work individually (Alomyan, 2004; Chen & Macredie, 2002; Chou, 2001).

In this research field, few studies inquire into the possible relationships that may exist between students’ stylistic characteristics and CL as a function of LA (Angeline, 2013; Angeli, Valanides, & Kirschner, 2009). Knowledge of these relationships help explain and understand the differences in LA in subjects when interacting with CBLE.

Cognitive load theory

Cognitive Load Theory (CLT) studies the existing relationship between working memory capacity and knowledge building that novices achieve when interacting in computational scenarios. In this manner, LA shall be affected if the structure and organization of the digital content overloads subjects’ limited memory resources. Consequently, the student is unable to effectively relate new information to the one stored in the long-term memory (Clark & Mayer, 2008; Sweller, 2006). Following this line of thought, it could be asserted that CL is all the mental activity imposed on the working memory when an individual is solving a learning task (Andrade-Lotero, 2012; Paas, Tuovinen, Tabbers, & van Gerven, 2003; Sweller, 2010).

CL is divided into three classes: (1) intrinsic, (2) extraneous, and (3) germane. With respect to Intrinsic Cognitive Load (ICL), it is inherent to the type of task to be developed. In other words, it considers the difficulty of domain knowledge to learn and the student’s prior knowledge. On the other hand, Extraneous Cognitive Load (ECL) is related to the information available in the computational environment that is irrelevant to task development and acts as a distraction that may divert the student’s attention. Finally, the Germane Cognitive Load (GCL) is directly responsible for knowledge building and is represented by actual LA (Andrade-Lotero, 2012; Sweller, 2006). The sum of the three loads is equal to total CL; thus, GCL will be favored when reducing both ICL and ECL (Chong, 2005; Sweller, 2006; Van Merriënboer & Sweller, 2005).

In CBLE, ICL cannot be manipulated by instructional design. However, in the design of computational environments, the objective is to reduce ECL and thus increase space in working memory to maximize GCL (Mayer & Moreno, 2003; Paas, Renkl, & Sweller, 2003; Sweller, 2010). Thereon, Medula (2012) found that ECL increases when audio, video, and text is articulated because of the overstimulation of the senses. In a more recent study, Andrade, Huang, and Bohn (2014) found that students exposed only to visual formats exhibited a lower ECL in comparison to those that combined visual and auditory information.

In another study, Cheon, Crooks, and Chung (2014), asked one group of students questions on content as they read texts (active segmentation), while another group was asked questions at the end of the interaction (passive segmentation). Results showed that in active segmentation students achieved better academic performances, probably when ECL decreased. More recently, Chen, and Wu (2015) reported a greater CL in students when interacting with videos that contain PowerPoint presentations and voice, compared to live conference and MOOC-type recordings.
Based on CLT, these studies contribute empirical evidence with respect to the use of some tools in the design of hypermedia environments that reduce CL. However, few studies focus on the use of scaffolding within the structure of web environments to favor students’ learning and reduce CL (Andrade-Lotero, 2012).

Metacognitive scaffolding

The concept of scaffolding was defined based on the Zone of Proximal Development (ZPD) posited by Vygotsky in his sociocultural theory of learning, which refers to the aid that an adult can give a child with the purpose of fulfilling the latter’s learning objectives (Tuckman, 2007; Wood, Bruner, & Ross, 1976; Wu & Pedersen, 2011). The scaffolding provides support to the student to successfully complete a learning task (Wood et al., 1976). Metacognitive scaffolding favor conscientious planning, monitoring, self-evaluation, and control of cognitive processes during learning task development in computational environments (Kim & Hannafin, 2011; Molenaar et al., 2010; Zhang & Quintana, 2012).

To that respect, Quintana et al. (2005) and Molennar et al. (2010) posit that metacognitive scaffolding are characterized for managing and regulating cognitive processes. This type of scaffolding is useful to the student to: (1) plan what they want to learn, in other words, it proposes defining learning goals, strategies, and timetables; (2) execute and monitor the progress of the proposed goals; and (3) reflect on the obtained results in order to review the effectiveness of the planning and adjust the strategies that have not been effective in achieving the learning goals. This process allows the student to acquire knowledge on how they learn, strategies to use, and time to invest according to the learning task.

Statement of the problem

In line with these statements, questions arise regarding the design of CBLE insofar as the use of scaffolding may be associated to the CL that the student experiences when interacting with these scenarios. Similarly, CL may be associated to the subject’s cognitive style. In this order of ideas, the present study posits the following research questions:

How does the metacognitive scaffolding influence CL and LA in students that learn in a hypermedia environment covering philosophy content? Do significant differences exist in CL between students with differing cognitive styles in the FDI dimension when they learn in a hypermedia environment?

Following this line of thought, the hypotheses that guide the present study are: (1) A reduction in CL exists and student performance increases because of the effect of the metacognitive scaffolding and (2) significant differences exist in CL between students with differing cognitive styles in the FDI dimension because of their stylistic differences.

Method

Design

The research was quasi-experimental with two eleventh (11th) grade groups from a public school of Bogotá – Colombia. The hypermedia environment is taken as the study’s independent variable with two values: group with metacognitive scaffolding and group without metacognitive scaffolding. The study’s dependent variables were: LA and CL. The latter with three values: ICL, ECL, and GCL.

Participants

This research was conducted with 54 eleventh grade students (26 women and 28 men) from a public school of the city of Bogotá – Colombia. The range of ages varied between 15 and 19 years ($M = 16.87, SD = 0.953$).
Instruments

Metacognitive scaffolding

The web-based learning environment on logic consists of three learning modules: (1) Definition and classification, (2) Aristotelian Logic: Parts, prepositions, and syllogisms, and (3) Symbolic Logic: Prepositions and classes, symbols and laws and connectors, as shown in Figure 1. The metacognitive scaffolding was implemented within its computational structure, which is displayed as pop-up windows. It is based on the model proposed by Winne and his colleagues (Hadwin & Winne, 2001). The scaffolding has the following characteristics:

Stage 1. Introduction to the learning task: The general content of each learning module is presented to the student, who is informed of schedules and spaces available to them for their development. In the first module, a general test on prior knowledge on philosophy was given to the student to get them to reflect on how much they know about the subject matter and on the learning strategies they could implement during the learning process.

Stage 2. Learning Planning: During this stage, the novice imposes on himself a learning goal based on his prior knowledge considering the following scale: (1) basic level; provides introductory and general information about the module’s subject matters; (2) intermediate level; examines in greater detail the content of each module, and (3) advanced level: in-depth study of each one of the subject matters. The scale’s objective is to consider their individual differences, as shown in Figure 2.

Subsequently, the novice establishes a work plan to achieve said goal. They set study times and choose the learning strategy according to the environment’s content and structure. This encourages reflection from the novice as a function of fulfilling the proposed goal.

Stage 3. Execution of the work plan: This stage begins with the implementation of the learning strategy. This activity has the objective of inducing the novice into metacognitive monitoring of the lessons learned, as shown
in Figure 3. According to this valuation, the student is in the capacity to make the necessary corrections regarding the self-imposed learning goal. This stage concurs with the metacognitive control process.

Stage 4. Learning Results: During this stage, the novice conducts the lesson’s final evaluation. They also reflect on the whole learning process. In other words, they evaluate the level of achievement reached, the planning of activities, and the chosen learning strategy to perform the corresponding adjustments in future learning modules, as shown in Figure 4.

Cognitive load questionnaire

The cognitive load questionnaire developed by Leppink, Paas, van Gog, van der Vleuten, and van Merriënboer (2014), which allows identifying students’ perception on CL through 13 items, was employed to determine students’ CL. For ICL, items 1 to 4. For ECL, items 5 to 8. Regarding GCL, items 9 to 13. The instrument is a self-reporting questionnaire and presents a Likert scale of 0 to 10, where 0 is “completely disagree” and 10 is “completely agree.” To obtain the grade in each load, the valuations were averaged. In this research, the instrument had a Cronbach’s alpha for ICL of 0.796; for ECL of 0.727, and finally, GCL of 0.816. Students answered the questionnaire three times, one for each lesson module.

Cognitive style test

EFT was used to determine cognitive style in the FDI dimension. The instrument proposed by Sawa (1966) consists of five subtests presented in separate pages. Each page has one simple figure and ten complex figures, which must be found within a given timespan. Previous applications of this test have shown an internal consistency varying between 0.85 and 0.9. (López-Vargas & Valencia-Vallejo, 2012; López-Vargas, Ibañez-
Ibáñez, & Chiguasuque-Bello, 2014). The EFT sample average was 24.72; standard deviation (SD = 8.886). Over a maximum grade of 50; the minimum value was 6 points and the maximum value was 41 points.

Students were grouped into FDs, intermediates, and FIs. This was done defining tertiles for the test’s total grade; hence, three grade ranges were identified: (a) 20 FD students (first tertile), (b) 16 intermediate students (second tertile), and (c) 18 FI students (third tertile).

**Learning achievement**

Students took three evaluations, one for each lesson module contained in the computational environment. All the evaluations consist of 10 multiple-choice points. The evaluations presented a high reliability of the instrument, Cronbach’s alpha was 0.872.

**Procedure**

To conduct the research, the educational institution’s board was contacted, who agreed to allow the eleventh-grade students’ participation in the study. Subsequently, students and teachers of philosophy were presented with the proposal. Then, parents were requested to give their informed consent regarding their children’s participation in the study, previously clarifying that the results would be confidential and for research purposes. Once the informed consents from all parents were gathered, the group was given the EFT.

Two computer labs were used to install the two hypermedia versions. One student course worked with the software version that contained the metacognitive scaffolding and a second course interacted with the software without the scaffolding. Students were assigned an identification password to access the software. The philosophy (logic) content was distributed in seven work sessions, each one with a duration of one hour per week. During each one of the work sessions, participants could not access Internet or other computer programs.

**Results**

**Effect of the metacognitive scaffolding on CL.**

A repeated measures ANOVA is used. There are two intra-subject variables: (1) CL with three values: intrinsic (I.L.), extraneous (E.L.), and germane (G.L.), in each one of the three lesson modules, and (2) LA in each one of the three modules. On the other hand, there are two inter-subject variables: (1) Hypermedia environment with two values: with metacognitive scaffolding and without scaffolding and (2) Cognitive scaffolding, with three values: field dependent, intermediate, and independent.

Table 1 shows a summary of the descriptive statistics of total CL in each one of the study modules of the students that worked in the hypermedia environment with and without scaffolding. Similarly, the novices’ cognitive style was considered.

<table>
<thead>
<tr>
<th>Software</th>
<th>Cognitive style</th>
<th>No.</th>
<th>Total load Module 1</th>
<th>Total load Module 2</th>
<th>Total load Module 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>With scaffolding</td>
<td>Field dependent</td>
<td>10</td>
<td>19.09(2.57)</td>
<td>20.76(1.28)</td>
<td>21.35(1.18)</td>
</tr>
<tr>
<td></td>
<td>Field intermediate</td>
<td>7</td>
<td>19.58(2.88)</td>
<td>21.59(0.64)</td>
<td>21.95(1.17)</td>
</tr>
<tr>
<td></td>
<td>Field independent</td>
<td>10</td>
<td>15.60(2.32)</td>
<td>19.89(1.62)</td>
<td>19.01(2.06)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>27</td>
<td>17.92(3.07)</td>
<td>20.65(1.43)</td>
<td>20.64(1.98)</td>
</tr>
<tr>
<td>Without scaffolding</td>
<td>Field dependent</td>
<td>10</td>
<td>22.56(3.85)</td>
<td>23.75(2.07)</td>
<td>23.59(2.44)</td>
</tr>
<tr>
<td></td>
<td>Field intermediate</td>
<td>9</td>
<td>21.74(2.85)</td>
<td>22.71(2.75)</td>
<td>22.29(2.41)</td>
</tr>
<tr>
<td></td>
<td>Field independent</td>
<td>8</td>
<td>22.06(4.14)</td>
<td>22.02(3.48)</td>
<td>20.84(3.24)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>27</td>
<td>22.14(3.52)</td>
<td>22.89(2.76)</td>
<td>22.34(2.82)</td>
</tr>
</tbody>
</table>

Firstly, total CL is obtained by adding the intrinsic, extraneous, and germane load. Mauchly’s test indicated that the sphericity assumption does not hold. The data show that the main effect of the total load variable is: ($X^2(2) = 30.54, p < .05$). Therefore, the degrees of liberty were corrected with Greenhouse-Geisser ($\varepsilon = .68$). Results show
a significant double interaction between total load and software \( (F(1.35, 64.96) = 6.62, p = .007, \eta^2 = .121) \). A significant difference also exists in the total load variable \( (F(1.35, 64.96) = 12.04, p < .001, \eta^2 = .200) \).

With respect to the inter-subject variables, significant differences exist between the type of software with CL \( (F(1.35, 64.96) = 20.73, p < .001, \eta^2 = .302) \) in favor of the subjects that interacted with the software version that included the metacognitive scaffolding. Similarly, significant differences exist between cognitive style with CL \( (F(1, 48) = 20.73, p < .001, \eta^2 = .302) \) in favor of the subjects that interacted with the software version that included the metacognitive scaffolding. Similarly, significant differences exist between cognitive style with CL \( (F(2, 48) = 5.04, p = .010, \eta^2 = .174) \). With respect to the use of the hypermedia environment with and without metacognitive scaffolding, the data show significant differences in total CL \( (F(1, 48) = 27.73, p < .001, \eta^2 = .312) \) in module 1. In module 2, significant differences also exist in total CL \( (F(1, 48) = 12.28, p < .001, \eta^2 = .204) \). Finally, in module 3, significant differences are found between the two student groups \( (F(1, 48) = 5.97, p = .018, \eta^2 = .111) \) as shown in Figure 5.

![Figure 5. Effect of hypermedia environment on total cognitive load in each module](image)

To conduct a more detailed analysis of each one of the CLs (intrinsic, extraneous, and germane) in each module, Table 2 shows the following descriptive data.

<table>
<thead>
<tr>
<th>Software</th>
<th>Cognitive style</th>
<th>No.</th>
<th>Module 1</th>
<th>Module 2</th>
<th>Module 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Scaffolding</td>
<td>Field Dependent</td>
<td>10</td>
<td>5.4 (1.35)</td>
<td>6.3 (1.38)</td>
<td>6.8 (1.37)</td>
</tr>
<tr>
<td></td>
<td>Field Intermediate</td>
<td>7</td>
<td>5.8 (1.59)</td>
<td>6.0 (0.99)</td>
<td>7.0 (0.96)</td>
</tr>
<tr>
<td></td>
<td>Field Independent</td>
<td>10</td>
<td>4.3 (0.81)</td>
<td>4.3 (0.92)</td>
<td>6.8 (0.94)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>27</td>
<td>5.1 (1.94)</td>
<td>5.4 (1.84)</td>
<td>7.3 (1.78)</td>
</tr>
<tr>
<td>Without Scaffolding</td>
<td>Field Dependent</td>
<td>10</td>
<td>7.4 (1.51)</td>
<td>7.3 (0.84)</td>
<td>8.0 (0.83)</td>
</tr>
<tr>
<td></td>
<td>Field Intermediate</td>
<td>9</td>
<td>7.2 (0.81)</td>
<td>7.0 (0.74)</td>
<td>7.7 (0.71)</td>
</tr>
<tr>
<td></td>
<td>Field Independent</td>
<td>8</td>
<td>7.7 (1.53)</td>
<td>7.6 (1.53)</td>
<td>7.3 (1.77)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>27</td>
<td>7.4 (1.29)</td>
<td>7.3 (1.25)</td>
<td>7.3 (1.28)</td>
</tr>
</tbody>
</table>

Mauchly’s test indicated that the sphericity assumption does not hold. The data show that the main effect of the module variable is: \( (X^2(2) = 30.54, p < .05) \) and of the CL variable is: \( (X^2(2) = 14.634, p < .05) \). The double interaction between module and CL yielded: \( (X^2(9) = 59.30, p < .05) \). Therefore, the degrees of freedom were corrected with Greenhouse-Geisser for module \( (\varepsilon = .68) \), CL \( (\varepsilon = .79) \), and for the interaction between module...
and CL (ε = .64). The results show a triple and significant interaction between module, CL, and software (F(2.56, 122.90) = 4.54, p = .007, η² = .086). Significant differences exist between the following double interactions: module and CL (F(2.56, 122.90) = 2.98, p = .042, η² = .058); between CL and software (F(1.58, 75.74) = 12.61, p < .001, η² = .208); and between module and software (F(1.35, 64.96) = 6.62, p = .007, η² = .121). Finally, significant differences exist in CL components (F(1.58, 75.74) = 12.28, p < .001, η² = .204) and module (F(1.35, 64.96) = 12.04, p < .001, η² = .200).

The multiple comparisons according to Bonferroni indicate that statistically significant differences exist (p < .05) in total CL, between FD (M = 7.28, SD = 1.5) and FI students (M = 6.63, SD = .16) and between intermediate (M = 7.21, SD = .17) and FI students (M = 6.63, SD = .16). No significant differences exist in total CL between FD and field intermediate subjects as shown in Figure 6.

Figure 6. Effect of cognitive style on total cognitive load

Regarding the student groups’ use of the hypermedia environment with and without scaffolding and the different CLs, the data show that in module 1, significant differences exist in ICL (F(1, 48) = 31.75, p < .001). Significant differences also exist in ECL (F(1, 48) = 22.70, p < .001) and no significant differences exist in GCL. In module 2, significant differences exist in ICL (F(1, 48) = 4.12, p = .048). Similarly, significant differences exist in ECL (F(1, 48) = 45.27, p < .001) and, as in Module 1, no significant differences exist in GCL. Finally, in Module 3, the same trend is identified; in other words, significant differences exist in ICL (F(1, 48) = 4.37, p = .042), in ECL (F(1, 48) = 14.85, p < .001), and no significant differences are present in GCL, as shown in Figure 7.

Figure 7. Effect of hypermedia environment on cognitive load (1 = intrinsic cognitive load, 2 = extraneous cognitive load, 3 = germane cognitive load)

Effect of the metacognitive scaffolding on learning achievement

A repeated measures ANOVA is used. The intra-subject variable is students’ performance in each one of the three lesson modules. There are two inter-subject variables: Hypermedia environment with three values: field
dependent, intermediate, and independent. Table 3 shows a summary of the descriptive statistics of LA in each one of the study modules considering the inter-subject variables.

### Table 3. Results of learning achievement in each module: Mean scores and standard deviations in parenthesis

<table>
<thead>
<tr>
<th>Software</th>
<th>Cognitive style</th>
<th>No.</th>
<th>Achievement Module 1</th>
<th>Achievement Module 2</th>
<th>Achievement Module 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>With scaffolding</td>
<td>Field dependent</td>
<td>10</td>
<td>76.40(5.74)</td>
<td>80.40(6.02)</td>
<td>77.50(8.89)</td>
</tr>
<tr>
<td></td>
<td>Field Intermediate</td>
<td>7</td>
<td>75.29(4.68)</td>
<td>82.29(4.15)</td>
<td>76.00(6.00)</td>
</tr>
<tr>
<td></td>
<td>Field Independent</td>
<td>10</td>
<td>71.80(7.66)</td>
<td>78.50(6.33)</td>
<td>75.80(6.92)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>27</td>
<td>74.41(6.41)</td>
<td>80.19(5.72)</td>
<td>76.48(7.27)</td>
</tr>
<tr>
<td>Without scaffolding</td>
<td>Field dependent</td>
<td>10</td>
<td>61.30(5.54)</td>
<td>66.60(5.91)</td>
<td>64.60(10.21)</td>
</tr>
<tr>
<td></td>
<td>Field Intermediate</td>
<td>9</td>
<td>59.56(7.57)</td>
<td>58.78(6.53)</td>
<td>59.78(10.40)</td>
</tr>
<tr>
<td></td>
<td>Field Independent</td>
<td>8</td>
<td>59.13(8.18)</td>
<td>63.88(4.29)</td>
<td>59.38(16.76)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>27</td>
<td>60.07(6.87)</td>
<td>63.19(6.45)</td>
<td>61.44(12.29)</td>
</tr>
</tbody>
</table>

Mauchly’s test indicated that the sphericity assumption does not hold. The data showed that the main effect of achievement is: \( \chi^2(2) = 80.09, p < 0.05 \). Consequently, the degrees of freedom were corrected with Greenhouse-Geisser (\( \varepsilon = .86 \)). Results show that significant differences exist only in the achievement variable \( F(1.73, 82.89) = 7.16, p = .002, \eta^2 = .130 \). With respect to the inter-subject variables, significant differences exist only between the hypermedia environment and LA \( F(1, 48) = 90.64, p < .001, \eta^2 = .654 \). Students that worked with the metacognitive scaffolding obtained a greater achievement \( M = 77.11, SD = 1.17 \) compared to students that did not use it \( M = 61.44, SD = 1.16 \), as shown in Figure 8. No significant differences were found in LA between students with different cognitive styles.

![Figure 8. Effect of hypermedia environment on learning achievement](image)

### Discussion and conclusions

#### First research question

The results showed that the implementation of the metacognitive scaffolding, within the structure of a hypermedia environment, positively influenced students’ LA.

The study’s analyses indicate that LA was significantly greater in students that used the metacognitive scaffolding in comparison to the achievement of their classmates that did not. Hence, in each lesson module the scaffolding favors metacognitive monitoring and process control insofar as the student learns to plan their study activities as a function of a self-imposed learning goal and to conduct a constant monitoring of their actual learning process, action they execute through different self-evaluations.
The self-evaluations carried out by the student leads them to reflect on the knowledge they have acquired throughout the modules. Accordingly, they can establish what concepts are still pending to be studied or reinforced to achieve the self-imposed learning goal. Thus, they can perform metacognitive control to review the content, change the learning goal, or adjust the learning strategy they have implemented to understand concepts and definitions. Hence, the scaffolding provides them with options and favors student’s autonomy, while the student undertakes the responsibility of their learning process.

In this sense, the study evidenced that the use of the metacognitive scaffolding fosters a more structured and systematic behavior in the student, which, probably, allows them to browse and perform different study activities in an organized fashion, adjust learning strategies, and process information in more detail with the objective of achieving a self-imposed goal.

With respect to CL, the results show that the metacognitive scaffolding reduced the perception of total CL in students that interacted with this hypermedia version. A separate analysis of each one of the CLs showed that the scaffolding significantly reduced ICL and ECL. The students’ perception of GCL was expected to decrease and enable a greater unloading of the working memory (Andrade-Lotero, 2012; Paas, Renkl, & Sweller, 2003); however, the scaffolding did not have any effect on the GCL.

It is noteworthy that the average GCL, in comparison to ICL and ECL, is greater and tends to be constant in the different lesson modules. This result is promising insofar as it is a first approach that indicates that metacognitive scaffolding may have positive effects on LA and, in addition, reduce ICL and ECL.

From the results, it is possible to assert that the students’ reflection on their own learning process can trigger them to change or adjust the chosen learning strategy, the manner of browsing, to read different content, etc. Thus, the use of the scaffolding helps the novice organize in a structured and systematic fashion their own learning process.

Because of the foregoing, it is possible to assert that the perception of the student’s ICL and ECL is lower in comparison to the group that did not use the scaffolding. These results evidence that the reduction in ICL, when beginning a learning task, helps the student easily learn the subject. Regarding ECL, it is possible that the scaffolding could be considered a distraction since it presents pop-up windows that aim to encourage the student to reflect on their own learning process; but, it is not so. The results show that ECL was reduced, which shows that the inclusion of scaffolding in hypermedia environments does not act as a distraction, but rather the scaffolding is a structural part of the computational environment.

On the other hand, no significant differences were found in GCL. A possible explanation for these results is that the perception of the two groups of students, with respect to content comprehension, in general, was highly valued, despite LA being significantly greater in the group that used the scaffolding.

This could indicate that the perception of the mental effort employed to comprehend the concepts and definitions covered in the hypermedia environment is always highly valued, independently of the LAs obtained in the evaluations the participants took at the end of each lesson module. It is noteworthy that the perception of learning difficulty was the same for both group of students. Thus, the scaffolding that was used did not exhibit any effect on GCL. It is likely that beginning a new learning process will always imply a high level of difficulty for novices and therefore germane load tends to be high. In accordance to the foregoing, it is suggested that this subject be studied in more detail insofar as other associate variables probably exist that influence these results. These results validate the first working hypothesis.

Second research question

With respect to the second research question, the results evidence that no differences exist in LA between students with different cognitive styles. It was verified that FD and field intermediate students achieve lessons equivalent to those obtained by their FI classmates. These results complement the findings of other studies that show that the use of computational scaffolding reduce individual differences in LA (López-Vargas, Hederich-Martínez, & Camargo-Uribe, 2012; López-Vargas & Valencia-Vallejo, 2012; López-Vargas & Triana-Vera, 2013).

With respect to total CL, the results show that significant differences exist between FI and FD students, and similarly, between FI and field intermediate students. No significant differences exist between FD and field
intermediate students. These results contradict the findings of Angeli (2013) and Angeli et al. (2009), who did not find significant differences between students with different cognitive styles in the FDI dimension with respect to CL. These contradictory results demonstrate the need for further research, especially when dealing with self-reporting instruments in which students tend to provide socially accepted answers and possess a high subjective component. Similarly, subject matter content and the type of software can have differential effects in students.

However, the results of the present study are coherent with the findings of Jia, Zhang, and Li (2014), who through contralateral delay activity, measure the brain’s electrical activity in visual working memory tasks in subjects with different cognitive styles in the FDI dimension. The study found that FI subjects have a greater capacity of isolating and filtering irrelevant elements in visual memory tasks in comparison to FD subjects. This situation could indicate that FI students have greater selective attention insofar as they more effectively inhibit the distractions present in computational scenarios and can probably process information in the working memory more efficiently.

Consequently, total CL employed in learning task development in FI students is smaller in comparison to their FD classmates, who are more prone to distractions and have a lower selective attention (Avolio, Alexander, Barrett, & Sterns, 1981; Hasher, Zacks, & May, 1999). The results of the present research may correspond with the efficient management of the working memory insofar as FI subjects are better at choosing relevant information and inhibiting distractions during learning task development in computational scenarios, situation that likely reduces CL. These results allow accepting the study’s second hypothesis.

In conclusion, studies suggest that the use of scaffolding may help reduce differences in LA in students with differing cognitive styles in the FDI dimension (López-Vargas, Hederich-Martínez, & Camargo-Uribe, 2012; López-Vargas & Valencia-Vallejo, 2012; López-Vargas & Triana-Vera, 2013) and favor CL reduction. Consequently, it is necessary to continue developing studies that provide empirical evidence in order for CBLE designers to develop more equitable and flexible computational scenarios, which favor subjects’ autonomy in learning and their performance in different levels of schooling.

Limitations and forecasts

The study was not an experimental-type research, considering that the groups that participated in the experience were previously constituted and were not randomly organized; therefore, the results cannot be generalized or extended to all students in the secondary educational system.

Another limitation refers to the initial characterization of the students to provide them differentiated support during the learning process based on their individual learning needs. In this sense, future research could consider including in the scaffolding elements that adapt to subjects’ stylistic characteristics to determine their effect on CL.

Considering that the instruments used so far to measure CL are based on self-reporting questionnaires, it is necessary to develop other indicators that allow objectively evidencing students’ cognitive effort.

References


López-Vargas, O., & TRIANA-VERA, S. (2013). Efecto de un activador computacional de autoeficacia sobre el logro de aprendizaje en estudiantes de diferente estilo cognitivo [Effect of a self-efficacy computational activator on the learning achievement in students of different cognitive style]. *Revista Colombiana de Educación, 64*, 225-244.


Improving Learning Analytics – Combining Observational and Self-Report Data on Student Learning

Robert A. Ellis1*, Feifei Han1 and Abelardo Pardo2

1Centre for Research on Learning and Innovation, Faculty of Education and Social Work, University of Sydney, Australia // 2School of Electrical Engineering, Faculty of Engineering and Information Technologies, University of Sydney, Australia // Robert.Ellis@sydney.edu.au // Feifei.Han@sydney.edu.au // Abelardo.Pardo@sydney.edu.au

ABSTRACT

The field of education technology is embracing a use of learning analytics to improve student experiences of learning. Along with exponential growth in this area is an increasing concern of the interpretability of the analytics from the student experience and what they can tell us about learning. This study offers a way to address some of the concerns of collecting and interpreting learning analytics to improve student learning by combining observational and self-report data. The results present two models for predicting student academic performance which suggest that a combination of both observational and self-report data explains a significantly higher variation in student outcomes. The results offer a way into discussing the quality of interpretations of learning analytics and their usefulness for helping to improve the student experience of learning and also suggest a pathway for future research into this area.

Keywords

Learning analytics, Self-report data, Observation, Student approaches to learning

Introduction

Using learning analytics as a tool to improve student learning has caught the imagination and research effort of much of the higher education sector (Siemens, 2013). Amongst a number of applications, it notably has been used to improve student success (Arnold, Hall, Street, Lafayette, & Pistilli, 2012; Martin et al., 2013), to better understand the nature of social learning amongst university students (Buckingham Shum & Ferguson, 2012), to improve approaches to learning design (Mor, Ferguson, & Wasson, 2015), and to guide university education strategy (Rientes et al., 2016).

Accompanying this growing use of learning analytics, there is serious debate about the extent to which they are useful as a tool for improving student learning (Lodge & Lewis, 2012; Lundie, 2014). One debate is about the objectivity of learning analytics; some argue that learning analytics are an objective measure of student activity, but others suggest that without understanding student intent behind the analytics, we have a poor context in which to interpret what the numbers mean (Boyd & Crawford, 2012). Another debate is that learning analytics tell us what students are doing when they learn in an online environment. Doubters argue that they only tell us what buttons they are clicking (Scheffel, Drachsler, Stoyanov, & Specht, 2014). A further debate surrounds the value of very large data sets. Some argue that the more analytics you have about student learning experiences the better, while others argue that a careful selection of analytics must be made in relation to the population sample, otherwise the additional metrics might just create noise in interpreting their meaning. As some studies suggest, indiscriminate approaches to the use of large datasets could lead to unintended consequences in learning interventions (Boyd & Crawford, 2012; Greller & Drachsler, 2012). To remedy some of the perceived shortfalls of learning analytics, some authors argue that the learning analytics should occupy a middle space, somewhere between learning theory and computational measurement, to improve the potential of learning analytics to really address concerns of the quality of student learning (Suthers & Veber, 2013). To achieve this, they recommend that additional analytic techniques accompany learning analytic procedures from such fields as epistemology and education studies.

To investigate methodological approaches to address some of the perceived shortfalls of learning analytics, this study investigates the first year experience of undergraduate engineering students in a blended course in two stages. In the first stage, it records their learning events in the online environment and analyses and interprets them in the context of their learning outcomes (Pardo, Han, & Ellis, 2016). While illuminating, this analysis alone could be left open to some of the criticisms described above. In the second stage, methodological approaches from Student Approaches to Learning (Pintrich, 2004) are used and the students’ response to closed ended questionnaires (Biggs, Kember, & Leung, 2001) about their experience of learning is investigated. The
outcomes of this analysis, when complemented by stage 1, both elucidates why some students are relatively more successful than others in the course and provides evidence which suggests why this might be the case.

The purpose of this study is to contribute to the international debate on the value of learning analytics for the quality of the student learning experience and how combined methodological approaches using observational and self-report evidence can improve our understanding of qualitative variation in student learning. By drawing on both types of data from the same experience of learning, this study is designed to see to what extent a combined use of the observational and self-report data improves our ability to use learning analytics to understand why some students are more successful than others.

Prior research

Student approaches to learning in higher education

Student approaches to learning (Pintrich, 2004), henceforth SAL, has systematically investigated university student learning of student learning and quality of learning outcomes. This framework has investigated how qualitative variation in students’ approaches to learning are closely related to prior experiences of learning, their perceptions of the learning context, student conceptions of learning and academic performance (e.g., Biggs & Tang, 2011; Marton & Säljö, 1976; Prosser & Trigwell; 1999; Ramsden, 2003). One of the important factors in learning processes is how students go about learning and whether their learning aims at rote memorization (i.e., surface approaches to learning) or their towards a meaningful understanding (i.e., deep approaches to learning) (Biggs & Tang, 2011). In one of the early seminal studies, Säljö (1979) reported an association between the qualitative differences in students approaches to learning and variations in students’ learning outcomes. He found that students who adopted a deep approach in a reading task were more likely to perceive it as intended by the author than those who approached it at a more surface level. Since then, a large number of studies have consistently identified variations of deep and surface approaches to learning and confirmed their association with learning outcomes in a wide variety of disciplines (e.g., Hay, 2007; Lindblom-Ylänne & Lonka, 1998; Lizzio, Wilson, & Simons, 2002; Prosser & Millar, 1989; Rossum & Schenk, 1984; Trigwell & Prosser, 1991; Trigwell & Sleet, 1990; Tang, 1998).

In the SAL framework, the approaches to learning adopted by a student are not a personal trait. They are instead related to factors described above (Entwistle, McCune, & Hounsell, 2003). Similar results have been found in blended learning contexts. Students with a fragmented conception of learning are more likely to approach face-to-face and online learning at more of surface level. On the other hand, those students that conceive learning as cohesive and integrated are more likely to adopt deep approaches in face-to-face and online learning (e.g., Bliuc, Ellis, Goodyear, & Piggott, 2010; Ellis, Goodyear, Calvo, & Prosser, 2008).

Learning analytics in higher education

In the last decade, the capacity of educational technology to record student interactions has led to the emergence of research areas such as Learning Analytics and Educational Data Mining (Johnson, Smith, Willis, Levine, & Haywood, 2011). The field is substantial and important to consider as a whole. The following highlights some of the key issues raised by the field relevant to this study.

There is an increasing number of learning analytic software systems that aim to use the records of students’ interactions to better understand their learning processes and environment (Baker & Siemens, 2014; Knight, Buckingham Shum, & Littleton, 2014; Lockyer, Heathcote, & Dawson, 2013). Learning analytic techniques use data mining techniques that process the rich data sets captured with technology to produce knowledge and improve the students’ learning experience. These systems have focused on aspects such as detecting students at risk to increase retention rates (e.g., Arnold et al., 2012), advising students on their career options (e.g., Bramucci & Gaston, 2012), analyzing curriculum structures (Méndez et al., 2014), and predicting academic performance (e.g., Antunes, 2010; Essa & Ayad, 2012a; Essa & Ayad, 2012b; Romero & Ventura, 2013).

Some studies have observed that the effect of using data mining and learning analytic methods from a purely technological perspective may have severe limitations (Buckingham Shum & Crick, 2012). To address potential and realized limitations, they suggest that techno-centric techniques should be combined with the findings in fields such as epistemology, educational studies, and pedagogy in order to better understand the links between learning and learning analytics (e.g., Buckingham Shum & Crick, 2012; Suthers & Vebert, 2013). Such an
approach can be referred to as the middle space of learning analytic research (Suthers & Vebert, 2013). Despite the identification of the benefits of this area of research, few studies can be located there.

One study that does begin to flesh out the middle space is a case study into university student learning in engineering (Pardo, Ellis, & Calvo, 2015). This study trialed the combination of instruments derived from SAL research and learning analytic techniques. In the study, several online tools recorded the interactions of students with various digital resources. Additionally, students’ conceptions and approaches to learning to three types of online learning activities (problem solving sequence, videos, and feedback), were collected through a qualitative survey. The mid-term examination results were used as students’ learning outcomes. The two sources of data were analyzed to examine their relationship academic results. The results showed that the frequency of interactions with three of the online learning tools were able to explain 25.98% of the variance in the mid-term exam results. Furthermore, deep approaches to problem sequences were related to higher marks, and surface approaches related to lower marks. The interpretation of results from both models translated into a set of actions to change the learning design and improve its quality.

This study contributes to the international debate on learning analytics and the benefits of combining methodologies to improve its usefulness. It seeks to confirm if the combination of SAL research and learning analytic techniques helps to identify meaningful qualitative differences in how students learned in an undergraduate engineering course, and to explore of a combination of the techniques improvise the quality of evidence for interpreting and predicting student learning success.

The main research question guiding the study is;
To what extent does the combination of learning analytic techniques and student approaches to learning methodologies improve our understanding of student learning success?

Method

Participants

The data for the study was collected from 291 undergraduate students enrolled in a first year engineering course in a large metropolitan Australian research-intensive university. Approximately 50% of the students agreed to participate in the study voluntarily, resulting in a sample of 145 students.

The learning context

The course is scheduled for the first year students in a Bachelor of Engineering Degree and has the following outcomes: (1) to design, build, configure, program, and test an electronic system for a specific engineering problem observing common professional practice; (2) to understand theoretical knowledge of how computers work, including concepts from the digital logic level to basic programming constructs; (3) to write reports about the design process and the results; and (4) to engage in team-based design work and creative tasks in order to solve an engineering problem. Apart from acquiring specific knowledge and skills in the content area, the course is also organized to build students’ generic attributes such as independent inquiry skills, communication skills, information literacy, intellectual autonomy, and the capacity to understand ethical, social, and professional issues.

The course adopts a typical blended learning approach during 13 weeks comprising both face-to-face and online work. The face-to-face component includes one weekly two-hour lecture, one weekly two-hour tutorial, and one weekly three-hour laboratory session. The online component is hosted in a custom-designed online environment integrated with a University Learning Management System (Blackboard.com). The online activities required students to interact weekly with digital material containing subject matter, visualize videos, interact with formative assessment elements, and submit summative assessments. Additionally, the system offered a dashboard with feedback about the individual participation rates in online activities.

Instruments

The data used in the study consisted of the student answers to a self-reported questionnaire about their learning experience, their use of the online learning tools as recorded by their online learning environment, and the
academic performance as provided by the final course mark. The choice of these data sets enabled the research design to consider the extent to which using both self-report data from the surveys and observational data from the learning analytics of the online learning environment informed their combined interpretation. A detailed description of each instrument is provided in the following sections.

The Revised Study Process Questionnaire

The Revised Study Process Questionnaire (R-SPQ) was used to collect the students’ self-report data (Biggs et al., 2001). The R-SPQ is a 20-item 5-point Likert questionnaire designed to evaluate the learning approaches of students. The theoretical context for its design is the Student Approaches to Learning framework (Biggs, 1987a; Biggs, 1987b; Biggs, 2011). In this study, the factor analysis indicated that the two-factor solution, deep and surface approach, fitted the empirical data, and the values of Cronbach’s alpha showed good reliability for both scales. Table 1 presents the outcomes of the analysis.

<table>
<thead>
<tr>
<th>Scales</th>
<th>Description of items</th>
<th>Rotated factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface approaches to study (.86)</td>
<td>3. My aim is to pass the course while doing as little work as possible.</td>
<td>.73</td>
</tr>
<tr>
<td></td>
<td>4. I only study seriously what’s given out in class or in the course outlines.</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>7. I do not find my course very interesting so I keep my work to the minimum.</td>
<td>.77</td>
</tr>
<tr>
<td></td>
<td>12. I generally restrict my study to what is specifically set as I think it is unnecessary to do anything extra.</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>15. I find it is not helpful to study topics in depth. It confuses and wastes time, when all you need is a passing acquaintance with topics.</td>
<td>.82</td>
</tr>
<tr>
<td></td>
<td>16. I believe that lecturers shouldn’t expect students to spend significant amounts of time studying material everyone knows won’t be examined.</td>
<td>.60</td>
</tr>
<tr>
<td></td>
<td>19. I see no point in learning material which is not likely to be in the examination.</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>20. I find the best way to pass examinations is to try to remember answers to likely questions.</td>
<td>.61</td>
</tr>
<tr>
<td>Deep approaches to study (.82)</td>
<td>1. I find that at times studying gives me a feeling of deep personal satisfaction.</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td>2. I find that I have to do enough work on a topic so that I can form my own conclusions before I am satisfied.</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td>6. I find most new topics interesting and often spend extra time trying to obtain more information about them.</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td>9. I find that studying academic topics can at times be as exciting as a good novel or movie.</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td>10. I test myself on important topics until I understand them completely.</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>13. I work hard at my studies because I find the material interesting.</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>14. I spend a lot of my free time finding out more about interesting topics which have been discussed in different classes.</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td>17. I come to most classes with questions in mind that I want answering.</td>
<td>.55</td>
</tr>
</tbody>
</table>

Events recorded in the online environment

As students engaged in this first year engineering course, they were expected to interact with a number of online activities which were made up of individual events (Pardo et al., 2015). The following events were recorded in the online environment:

- Duration: student time spent working on an activity.
• Dashboard: student views of a dashboard with information about the engagement with the weekly activities
• Col-Exp: student views of various sections of the course notes
• Resource: student views of any page of the course notes
• Video: interaction with a video (video is loaded, played, paused or finished)
• MCQ: any interaction with a multiple-choice question embedded in the course notes
• VMCQ: any interaction with multiple-choice questions placed next to a video
• Exercise: student answers to summative assessment exercises.

For each type of event and student, eight variables were calculated with the accumulated number of events over the semester. The descriptive statistics of these variables are shown in Table 2.

Table 2. Descriptive statistics of frequencies of online learning events

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>0</td>
<td>93</td>
<td>11.52</td>
<td>17.45</td>
</tr>
<tr>
<td>Dashboard</td>
<td>0</td>
<td>233</td>
<td>31.10</td>
<td>41.84</td>
</tr>
<tr>
<td>Col-Exp</td>
<td>59</td>
<td>1182</td>
<td>421.97</td>
<td>234.36</td>
</tr>
<tr>
<td>Resource</td>
<td>138</td>
<td>2492</td>
<td>818.07</td>
<td>443.00</td>
</tr>
<tr>
<td>Video</td>
<td>0</td>
<td>2890</td>
<td>338.59</td>
<td>395.48</td>
</tr>
<tr>
<td>MCQ</td>
<td>0</td>
<td>3054</td>
<td>233.01</td>
<td>300.50</td>
</tr>
<tr>
<td>VMCQ</td>
<td>0</td>
<td>5598</td>
<td>191.05</td>
<td>471.17</td>
</tr>
<tr>
<td>Exercise</td>
<td>353</td>
<td>9957</td>
<td>2723.49</td>
<td>1419.81</td>
</tr>
</tbody>
</table>

*Notes. Col-Exp = collapse and expand, MCQ = multiple choice questions, VMCQ = multiple choice questions embedded in videos.*

The frequencies show quite large difference between event types. For example, some variables have a range of values that start at zero (Duration, Dashboard, Video, MCQ, and VMCQ) whereas the minimum value for the Exercise variable is 353. The same effect can be observed with the standard deviations (SDs) reflecting the distinct use students make of the online tools. Due to these large differences, all variables were standardized (mean equal to 0 and a SD of 1) to facilitate comparisons among them.

**Academic performance**

The information about academic performance was collected using the final course mark. The final mark was calculated by aggregating six types of assessment tasks: exercises to prepare the lecture (10%), exercises to prepare the tutorial (10%), one written report about a laboratory session (5%), a written report, presentation and demonstration of a collaborative project (15%), a midterm exam (20%) and the final exam (40%). The potential value range for this variable is 0 to 100, but the marks for the participants ranged from 20 to 98.50, with a mean of 65.50, and a SD of 16.12. The large SD of this variable indicates a wide spread of final marks in the course. As in the case of the previous variables and to facilitate interpretation of the interactions among variables, the final mark was also transformed into a z-score with mean 0 and a SD of 1.

**Data collection procedure**

Prior to the data collection, the study was reviewed and approved by the institutional ethics committee. Students were informed that their participation in the study was voluntary, about the use of the online environment to monitor their interaction, that the information collected would be anonymized, and used only for research purpose. Written consent was obtained to use this data together with their course results. The self-report data from the R-SPQ questionnaire was collected in class towards the end of the semester so that students could reflect on the learning processes of the whole course. The data from the online environment was collected throughout the entire course.

**Data analysis**

The data analysis was performed in three stages; identifying qualitative differences in the student approaches to learning (deep and surface variables), using correlation and cluster analysis to identify the strength of associations amongst the approach variables, the online learning analytics of the events and academic
achievement; and finally hierarchical regression analysis to identify which variables most explained variance in the student experience.

In the first stage, a series of Exploratory Factor Analysis (EFA) using Principal Component procedure followed by varimax rotation were used to examine the factor structure of the answers to R-SPQ (self-report data). A number of criteria were used to determine the number of scales and corresponding items. The scree plot was used to determine the possible numbers of solutions for the scales. We deleted those items whose coefficients were < .40 within a factor, and those with high multiple coefficients loaded across factors (Field, 2013). The internal consistency of the retained scales was verified using Cronbach’s alpha reliability analyses (alphas above .70).

In the second stage, we investigated the relationship between the three data sources; the self-report data, the data collected from the students’ engagement with the eight learning events in the online environment, and their academic performance data as measured by their course mark. Two analyses were used to explore these relations in order to increase the overall integrity of the approach (Prosser, Ramsden, Trigwell, & Martin, 2003). First, we used correlation analysis to show the interrelationship between pairs of variables. Second, we used a hierarchical cluster analysis using the self-report data and academic performance to identify subgroups of students with similar learning experiences within the population sample. The cluster membership information derived from the previous step was used to perform one-way ANOVA to see whether students in different clusters differ from each other with respect to their academic performance.

In the third stage of analysis, we used hierarchical multiple regression analysis to examine the relation between the combined self-report and learning event data and the course marks of the students. The regression model was built with those variables for which a significant correlation with academic outcomes was established by bivariate correlation. The dependent variable of the model was the students’ course mark. The independent variables were: the surface approaches to study (from the self-report data), the number of Dashboard learning events, number of Collapse-and-Expand events which indicate student views of different areas of the online environment, number of Resource events which are student views of the course notes, number of Multiple Choice Question events, and number of Exercise events involving summative assessment exercises. All these variables are from the online environment.

Two hierarchical multiple regression models were obtained for comparison purposes, one with self-report data only and a second which included the observational data. The first one was produced considering only the surface approaches to study from the self-report data as there is an established body of research indicating its suitability to predict learning outcomes (e.g., Trigwell, Ashwin, & Millan, 2013; Trigwell, Ellis, & Han, 2012). The second regression model was calculated by adding the five variables obtained from the online environment data. The two models were compared using Cohen’s $f^2$ (Cohen, 1992).

**Results**

**Exploratory factor analysis and reliability of the scale**

The results of the first stage of analysis confirm the reliability and validity of the approach scales used in all the analyses. Sixteen out of the twenty items were retained for the two-factor solution with eight items in each factor, shown in Table 1. The two factors are the deep and surface approaches to study. The Eigen-values of the surface and deep factors were 4.26 and 3.69, explaining 26.59 % and 23.09% of the total variance respectively. The reliability analyses showed that the surface and deep approaches to study scales had Cronbach’s alpha of .86 and .82, confirming their reliability.

**Correlation and cluster analyses**

Stage 2 results include the correlation and cluster analyses. The correlation results identify the strength of associations amongst the variables and the cluster analyses identify the distribution of those associations amongst sub-groups in the population sample who report similar experiences. The correlations between the deep and surface approaches to study (derived from the self-report data), the number of interaction with online tools (observational data), and their academic performance are presented in Table 3. The deep approach to study does not show a statistically significant relation to academic performance. However, it has a weak positive correlation with the number of Duration events ($r = .17, p < .05$), the number of Multiple Choice Question events ($r = .18, p
< .05), and the number of Exercise events was \( r = .21, p < .05 \). The associations between the surface approach to study and different types of online events were not significant, but we found that the surface approach to study significantly and negatively correlated with the final course mark \( r = -.31, p < .01 \). The results also show that the final marks have strong and significant correlations with almost all the event counts in the online environment. These positive correlations suggest that the more frequently a student engaged with the online environment the higher final course marks they tended to obtain.

**Table 3. Correlation analysis**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Deep approaches to study</th>
<th>Surface approaches to study</th>
<th>Course mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface</td>
<td>-.08</td>
<td>.05</td>
<td>-.31**</td>
</tr>
<tr>
<td>Duration</td>
<td>.17*</td>
<td>-.12</td>
<td>.11</td>
</tr>
<tr>
<td>Dashboard</td>
<td>.08</td>
<td>-.10</td>
<td>.24**</td>
</tr>
<tr>
<td>Col-Exp</td>
<td>.12</td>
<td>.08</td>
<td>.35**</td>
</tr>
<tr>
<td>Resource</td>
<td>.14</td>
<td>.04</td>
<td>.43**</td>
</tr>
<tr>
<td>Video</td>
<td>.11</td>
<td>.01</td>
<td>.14</td>
</tr>
<tr>
<td>MCQ</td>
<td>.18*</td>
<td>.07</td>
<td>.28**</td>
</tr>
<tr>
<td>VMCQ</td>
<td>.12</td>
<td>-.14</td>
<td>.14</td>
</tr>
<tr>
<td>Exercise</td>
<td>.21*</td>
<td>-.10</td>
<td>.38**</td>
</tr>
</tbody>
</table>
| Notes. **p < .01; *p < .05. Col-Exp = collapse and expand, MCQ = multiple choice questions, VMCQ = multiple choice questions embedded in videos.**

A hierarchical cluster analysis using Ward’s method was conducted using the two factors of the students’ learning experience (self-report data) and academic performance to identify and characterize subgroups of students with similar features. Based on the value of the squared Euclidean distance between clusters, a two-cluster solution was obtained. The results are shown in Table 4. The second and third columns from the right show mean and SD for the 11 variables. The three columns on the left of the table show the results of ANOVA.

**Table 4. Summary statistics of the two-cluster solution**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Deep cluster (43) Mean (SD)</th>
<th>Surface cluster (102) Mean (SD)</th>
<th>F</th>
<th>p</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep</td>
<td>0.93 (0.73)</td>
<td>-0.39 (0.83)</td>
<td>82.11</td>
<td>.00</td>
<td>.37</td>
</tr>
<tr>
<td>Surface</td>
<td>-0.52 (1.18)</td>
<td>0.22 (0.82)</td>
<td>18.74</td>
<td>.00</td>
<td>.12</td>
</tr>
<tr>
<td>CM</td>
<td>0.53 (0.99)</td>
<td>-0.23 (0.82)</td>
<td>19.77</td>
<td>.00</td>
<td>.12</td>
</tr>
<tr>
<td>Duration</td>
<td>0.20 (1.09)</td>
<td>-0.08 (0.95)</td>
<td>2.42</td>
<td>.12</td>
<td>.02</td>
</tr>
<tr>
<td>Dashboard view</td>
<td>0.32 (1.20)</td>
<td>-0.14 (0.87)</td>
<td>6.69</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>Col-Exp</td>
<td>0.27 (1.02)</td>
<td>-0.11 (0.97)</td>
<td>4.46</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td>Resource</td>
<td>0.33 (1.08)</td>
<td>-0.14 (0.93)</td>
<td>7.10</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>Video</td>
<td>0.26 (1.39)</td>
<td>-0.11 (0.76)</td>
<td>4.19</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td>MCQ</td>
<td>0.61 (1.54)</td>
<td>-0.15 (0.35)</td>
<td>7.57</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>VMCQ</td>
<td>0.27 (1.77)</td>
<td>-0.11 (0.28)</td>
<td>4.72</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td>Exercise</td>
<td>0.44 (1.25)</td>
<td>-0.18 (0.81)</td>
<td>12.58</td>
<td>.00</td>
<td>.08</td>
</tr>
</tbody>
</table>

Notes. Deep = deep approaches to study, Surface = surface approaches to study, CM = course mark, Col-Exp = collapse and expand, MCQ = multiple choice questions, VMCQ = multiple-choice questions in videos.

Table 4 shows the students were classified into two clusters: 43 students were assigned to the “Deep” cluster and 102 students were assigned to the “Surface” cluster. The results of one-way ANOVA confirmed our hypothesis that the self-report data and the academic performance had statistically significant contrasts between the two clusters of students with values of \( F(1, 144) = 82.11, p < .01, \eta^2 = .37 \), and \( F(1, 144) = 18.74, p < .01, \eta^2 = .12 \), for the deep and surface approaches respectively. Statistically significant differences were also found between the two clusters on all variables in our observational data although with small values of Eta squared (effect size).

Table 4 shows the students were classified into two clusters: 43 students were assigned to the “Deep” cluster and 102 students were assigned to the “Surface” cluster. The results of one-way ANOVA confirmed our hypothesis that the self-report data and the academic performance had statistically significant contrasts between the two clusters of students with values of \( F(1, 144) = 82.11, p < .01, \eta^2 = .37 \), and \( F(1, 144) = 18.74, p < .01, \eta^2 = .12 \), for the deep and surface approaches respectively. Statistically significant differences were also found between the two clusters on all variables in our observational data although with small values of Eta squared (effect size).

The mean and SD of the self-report variables in the resulting clusters are consistent with their characterization. The deep cluster has a relatively higher rating for the deep approach (mean = 0.93) and relatively lower rating for the surface approach (mean = -0.52), and the situation is reversed for the surface cluster (mean = -0.39 and mean = 0.22 for the deep and surface approach respectively). Furthermore, the variables derived from the observational data have systematically higher means for the deep cluster than the surface cluster. Deep learners engaged more frequently with the various online resources than those in the surface cluster. The final mark variable also shows a pattern consistent with the self-report and observational data. There is a significant
difference between the score of the students in the deep cluster (mean = 0.53) and the surface cluster (mean = -0.23). The main conclusion from this result is that both self-report and observational variables are important when identifying the relations with the overall academic performance.

Multiple regression analysis

The results of the third stage of analyses compare two linear models of multiple regression to see the effect of combining the self-report and observational variables that had significant correlation with the final score (see Table 5).

<table>
<thead>
<tr>
<th>Variables</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>Adjusted R²</th>
<th>p</th>
<th>f²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface</td>
<td>-0.31</td>
<td>0.08</td>
<td>-0.31 **</td>
<td>-3.84</td>
<td>0.09</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td>Dashboard</td>
<td>0.03</td>
<td>0.08</td>
<td>0.03</td>
<td>0.31</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col-Exp</td>
<td>-0.34</td>
<td>0.19</td>
<td>-0.34</td>
<td>-1.84</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource</td>
<td>0.88</td>
<td>0.21</td>
<td>0.88 **</td>
<td>4.28</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCQ</td>
<td>0.29</td>
<td>0.12</td>
<td>0.29 *</td>
<td>2.40</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exercise</td>
<td>-0.26</td>
<td>0.18</td>
<td>-0.26</td>
<td>-1.40</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. **p < .01; *p < .05. Surface = surface approaches to study, Col-Exp = collapse and expand, MCQ = multiple choice questions.

Before performing the multiple regression analysis, a series of tests were conducted to examine the assumptions that may affect its reliability. The initial analysis of standard residuals confirmed the absence of outliers (Std. Residual Min = -2.15, Std. Residual Max = 2.54). The tolerance values were above 0.10 for all the variables confirming the absence of multicolinearity. Finally, the Durbin-Watson statistic verified the absence of autocorrelation (Durbin-Watson = 2.08).

Model 1 included only the self-reported surface approach to learning whereas model 2 was obtained with the self-report variable and the five observational variables which have significant correlations with the final score. The results for both models are shown in Table 5. Model one reveals that the surface approach to study contributed significantly to the regression model: F(1, 143) = 14.72, p < .01, f² = .10. However, the model accounted for only 9% of the variation in the final marks. Model 2 also returned a significant result: F(5,138) = 10.76, p < .01, f² = .52, however in this case, the model accounts for 34% of the variation. This means an additional 25% of variation in students’ academic performance is explained when the observational variables are included in the model. The increase in R² also means that the prediction intervals obtained with the model using all six variables (self-report and observational) will be significantly smaller.

A more detailed analysis of this result shows that three independent variables significantly predicted students’ academic performance: the surface approach to study (β = -.29, p < .01), the number of times an online resource was accessed (β = .88, p < .01), and the number of multiple-choice questions answered (β = .29, p < .05). The remaining variables, Dashboard, Col-Exp, and Exercises did not make significant predictions to the final course marks. These results show that the combination of self-report and observational data provides a better predictive model of students’ learning outcomes.

Discussion

This study combined research methodologies from SAL research and learning analytics to examine the relationship between students’ learning experience, their interactions with online learning tools, and their learning outcomes. To examine the relationship among variables, we conducted correlation analysis, cluster analysis, and multiple regression. The first one looked into the pairwise relationship between variables. The second identified sub-groups of students within the population sample which reported qualitatively different experiences of learning. The third analysis used hierarchical regression analysis and showed that both the learning experience reported by students and the behavior recorded while interacting in an online environment significantly explained the final course marks. Together, the two sources of the data could predict approximately
one third of the variance, with the interactions of online learning tools accounting for more than double of the variance (25%) than students’ reporting their learning approaches (9%).

Consistent with previous research in SAL, our results confirmed that students who adopted a surface approach to study tended to associate with poorer performance in the same course. Our study showed how students who reported using a deep approach to learning also tended to interact more frequently with the online environment. We also found a positive relation between the frequency of student interactions with the online learning tools and the final course mark. Out of the eight tools deployed for the course, five of them had positive relationship with the final course marks. Our results were consistent with those presented by Romero-Zaldívar, Pardo, Burgos, and Kloos (2012) in which interaction counts of a number of online tools recorded by a virtual appliance were positively related to students’ final grades among second year engineering students (correlation ranged between .07 and .16). However, the study described in this paper shows stronger associations (correlation ranged between .24 to .43).

The relationship between students’ approach to learning, interaction with online tools, and final course marks were further substantiated using cluster analysis. Students in the deep cluster reported using a deep rather than a surface approach to learning, and obtained relatively better course marks. Significantly, this cluster was observed to have a relatively more intense interaction with the online learning events. These were found to be significantly more frequent among students in the deep cluster than those in the surface cluster. These results support the claim that the observed student behavior is consistent with what they reported in the R-SPQ questionnaire. While the study described by Purdo et al. (2015) found qualitatively differences on how students approached different online learning events, this study suggests that the amount of interaction online can also contribute to students’ achievement in a course. We speculate that both the quantity and the quality of using online tools helps to explain variations in the learning outcomes, and the nature of these associations needs to be empirically examined in increasingly fined-grained analysis in future studies.

The results of the hierarchical regression analysis indicated that two of the five learning tools, were significant predictors of academic performance when combined with the surface approach to learning. We find this to be a key outcome of the study and it provides a response to the research question which motivated the study. It highlights the value of combining the different types of data sources (self-report and observational) when assessing the learning experience of university students in order to provide evidence for improving outcomes.

The results offer some key theoretical implications. A key aspect of student learning research theory is the close association between the quality of approaches to learning and relatively higher academic achievement. The results of this study suggest that this is the case, however not just with the approaches and academic achievement, but also with the students’ use of the learning technologies themselves. This result highlights the relational contribution of “material” elements of the experience to learning outcomes and offers an important area for future investigations and theory building about the contributions of non-human elements in learning (Fenwick, 2015). Material elements of the student experience are not typically investigated in student learning research. This outcome calls for further work into the role of material elements in the student experience, particularly the interplay between approaches and learning technologies for example.

Conclusions

The study described in this paper is an initial effort to combine elements of SAL theory and learning analytics in an investigation into a blended university course. The results have revealed how the insight and understanding of the student learning experience can be improved by combining instruments that capture self-reported data, and observational data. While each data source can be used separately, this study has offered a quantified improvement through their combined use and provides one way of fleshing out the middle space between learning and analytics (Suthers & Vebert, 2013).

Various interesting avenues have emerged in this study that warrant further exploration. While our study adopted the theoretical framework derived from SAL, it is possible to explore the integration of other theories in educational psychology and learning analytics. For instance, we can examine the relationship between students’ self-efficacy, motivation, and experienced emotions, and their behavior in a learning environment. Second, the variables in the study derived from the observational data were summarized as event counts. A more detailed type of indicator could be derived from this data if combined with descriptions of students’ conceptions of using different online tools. Finally, experiences of learning and teaching are highly complex interactions shaped by a large number of interdependent factors. The results of this study offer sufficient promise to continue to
investigate the attenuated approaches to the collection and analysis of learning analytics so that researchers and the stakeholders of such research receive a stronger context in which to evaluate the meaning of the results.

Acknowledgements

The authors are pleased to acknowledge the financial support of the Australian Research Council through grant DP150104163.

References


Metacognitive Support Accelerates Computer Assisted Learning for Novice Programmers

Siti Nurulain Mohd Rum1* and Maizatul Akmar Ismail2

1Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Selangor, Malaysia // 2Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia // snurulain@upm.edu.my // maizatul@um.edu.my

*Corresponding author

(Submitted February 29, 2016; Revised May 20, 2016; Accepted June 16, 2016)

ABSTRACT

Computer programming is a part of the curriculum in computer science education, and high drop rates for this subject are a universal problem. Development of metacognitive skills, including the conceptual framework provided by socio-cognitive theories that afford reflective thinking, such as actively monitoring, evaluating, and modifying one’s thinking, has been identified as important for novice programmers. Studies have shown that metacognitive skills can be nurtured through the use of technology blended into educational activities. Designing metacognitive-related activities that focus on both social and cognitive development is both theoretically and practically challenging, especially in supporting the teaching and learning of computer programming. This paper describes six commonly-used strategies, viz., metacognitive scaffolding, reflective prompts, self-assessment, self-questioning, self-directed learning and graphic organizers, identified as important features that can be incorporated into computer-assisted learning tools in supporting computer programming learning. An experimental study was conducted to determine the effectiveness of these strategies. The results show that they helped learners by improving their performance in learning computer programming.

Keywords
Metacognitive, Support system, Self-regulation, Reflective prompts, Self-assessment, Self-questioning, Computer programming

Introduction

Interactive online educational technologies such as intelligent tutoring systems (ITS) can provide a good platform for performing research related to metacognition. Assessment of fine-grained tracking of students’ cognitive abilities and their metacognitive behaviors can be provided by such systems. Designing a metacognitive support system focused on both cognitive and metacognitive development can be very challenging. The term metacognitive is most often associated with (Flavell, 1979) who has defined metacognitive knowledge as the process of acquiring knowledge about cognition to control cognitive processes. The role of metacognition in learning to solve computer programming problems is very important. Metacognitive management strategies are more often used by outstanding programming students than by lower-performing students (Bergin, Reilly, & Traynor, 2005). In fact, the more complex a programming problem, the greater the need for metacognitive control, purposeful reflection, and positive feedback (Havenga, 2011). A programmer must apply in-depth reading skills and meta-comprehension to judge how clearly and effectively he or she understands a programming problem. Furthermore, programmers must be skilled in problem-solving processes, apply appropriate programming approaches, and be able to correct programming errors, as well being able to think deeply about their programming solutions and test program output. Such problem-solving steps require metacognitive control such as planning (plan the solution), monitoring (monitor the design and development of the program) and evaluation (test and reflect on the programming solution). A well-trained programmer is someone equipped with good analytical thinking and problem-solving skills (Soloway & Spohrer, 2013). Students should therefore manage their skill with respect to programming processes, motivate their decisions, articulate their actions, and investigate alternative solutions to improve the quality of their programs. The teacher has a responsibility to support students in developing metacognitive skills and applying them during program development. Although there are methods of instruction (i.e., cognitive approach, motivation approach) the most valuable and effective methods involve a combination of theory and practice (Li, Zhang, Du, Zhu, & Li, 2015). The underlying strategies and knowledge of cognitive processes must be given to a learner along with opportunities to practice and apply both metacognitive and cognitive strategies. For development of metacognitive regulation, it is also important to evaluate the outcome of their efforts (Wegener, Silva, Petty, & Garcia-Marques, 2012). Yet, to date, there are many attempts have been made to provide the support learning tool for novice to learn computer programming (Sorva, Karavirta, & Malmi, 2013; Verdú et al., 2012). However, finding similar works that specifically discuss on improving novices’ metacognitive skills using support learning tool could not be located. In the next section, we discuss the characteristics of a metacognitive support system.
for learning computer programming; they are identified as scaffolding, reflective prompts, self-assessment, self-questioning, self-directed strategies and graphical organizer.

**Metacognitive scaffolding**

Difficulties encountered while learning computer programming are a universal problem. There have been numerous attempts to address these difficulties (e.g., (Apiola, Tedre, & Oroma, 2011; M. Rum, Nurulain, & Ismail, 2014; S. N. M. Rum & Ismail, 2014; Soloway & Spohrer, 2013), but challenges still remain, so an optimal support mechanism using learner and effective instructional strategies should be developed to provide an optimal learning environment for learning computer programming. These difficulties show that some programming skills required by novice learners may be beyond their capabilities. Scaffolding is a critical component in facilitating students’ aptitude for programming (Bickhard, 2013; Feyzi-Behnagh et al., 2014). Scaffolding entails providing students with assistance that would increase their competence on an as-needed basis (Kim & Hannafin, 2011). With scaffolding, learners can engage in activities that otherwise would be beyond their abilities. Scaffolding refers to a variety of instructional techniques used by computer tutors and humans as well as such pedagogical agents as tools, guides and strategies in helping learners to develop understandings and competency that might otherwise be beyond their grasp (D'mello & Graesser, 2012). Learning difficulties can be reduced by providing such support outside the classroom, and it should augment, not replace, classroom learning. Several studies have shown that students often fail to gain basic understanding and exhibit poor performance when they use a computer-based learning environment to learn about complex topics without scaffolding (Azvedo, Cromley, Winters, Moos, & Greene, 2005; Greene & Land, 2000). Researchers have therefore begun emphasizing the significance of procedural, metacognitive, embedded, strategic, and conceptual scaffolds in computer-assisted learning environments. A review by (Jumaat & Tasir, 2014) has categorized scaffolding used in online learning into four main types; (1) conceptual scaffolding, (2) procedural scaffolding, (3) strategic scaffolding and (4) metacognitive scaffolding. Conceptual scaffolding guide learners to the key concepts of learning, procedural learning helps learners use appropriate resources as well as tools effectively; strategic scaffolding helps learners find appropriate strategies and methods in solving complex problems and metacognitive scaffolding helps learners to be prompted and reflected about what they are learning throughout the learning process. Metacognitive scaffolding not only promotes higher order thinking of learners but it also support learners ability to plan ahead.

**Reflective prompts**

Problem-solving is a common activity used as a teaching and learning approach in computer programming education because it helps students to develop different cognitive abilities. Solving problems requires a great deal of cognitive effort in activities such as finding solutions, identifying problems, and testing hypotheses. Prompting using self-regulated support and learning through problem solving is recognized as a valid instructional approach, especially in computer-assisted learning fields where simple prompting can be easily implemented (Bannert & Reimann, 2009, 2012). Research specifically focusing on self-regulated learning has shown that learners face many difficulties in performing self-regulation activities, and they may process information spontaneously, often failing to achieve desired learning outcomes (Bannert, 2006; Bannert & Reimann, 2012). Prompting students while engaged in task performance is an instructional strategy that seems promising for carrying out specific regulated activities. This method can be employed for supporting and guiding a learner in being self-regulated during the problem-solving process (Bannert & Reimann, 2012). Question prompts have been found by (Davis, 2000; Kim & Hannafin, 2011; Lai, 2008) to be an effective approach for eliciting reflection and helping student focus their attention as well as to monitor their learning through elaboration on the questions asked. To strength the quality of reflection, (J. A. Moon, 2004) proposed structuring reflection with questions. By performing reflective activities, learners would be able to identify the importance of various activities and thereby be able to develop understanding in a larger context such as responding to question prompts (Amulya, 2004). In a peer-tutoring context, recent research has shown evidence of a positive influence of an instructor’s question prompts on a learner’s reflective learning. Research shows that an instructor’s question can motivate learner metacognition that could positively influence learning opportunities (Wu & Looi, 2011). The importance of incorporating question prompts into Intelligent Learning Environments (ILE) designs have also been recognized by researchers in ILE (e.g., (Murray et al., 2013; Van der Meij & de Jong, 2011). Question prompts have been used to scaffold learners towards appropriate learning and positive evidence of effectiveness of this approach in helping learners have been found, for example, in application to ill-structured problem-solving (Xie & Bradshaw, 2008) and integration of knowledge (Davis, 2000). Students can reflect on their own thoughts by motivating and activating them to analyze and think about the effectiveness of
strategies chosen to help them oversee, regulate, and control the application of procedures in a specific situation. Whether to display the question prompt in ILE usually depends on the purpose and intention of a specific interposition. A learner should receive the prompt from the learning tool at an appropriate time, e.g., at the moment where assistance is required; presenting a prompt in an inappropriate manner or at an inappropriate time will only cause cognitive overload (Ifenthaler, 2012; Thillmann, Künsting, Wirth, & Leutner, 2009). Generally, the assistance would be provided during, before, or after the learning sequence presentation. The presentation of a prompt during the sequence of learning is reasonable if the intention is to activate a learner’s monitoring skill in problem-solving activities, but if the objective is to motivate the learner to evaluate certain problem-solving activities, then presentation after the learning sequence would be better. If one wishes to trigger learner reflection on determining an approach to solving a problem it would be appropriate to present a prompt before presenting the problem-solving sequence. Another crucial aspect is how a prompt can best be embedded to provide the learner with an optimal scaffold. There are two categories of reflective prompts, generic and directed; a generic prompt would seem more effective than a directed prompt because it gives the learner autonomy (Davis, 2003). On the other hand, a directed prompt would generally helpful for novices who are lack of problem-solving skills (Davis, 2003).

**Self-assessment**

One issue in teaching computer programming is student motivation. The learning process can be both effective and meaningful if students are given an opportunity to understand objectives for achieving goals and following their own learning processes. Focus on learning outcomes has been dominant at many universities (J. Moon, 2002). Learning outcomes refer to the knowledge level of learners in terms of what they know and are expected to do as a result of learning activity so this information can be used for goal-setting (Anderson et al., 2001). Self-assessment is a common method for assessment. Research has revealed that students perform better if they are able to evaluate their own work or the work of others (Anderson et al., 2001; Cassidy, 2006). Students who oversee and accurately evaluate their own performance may respond with appropriate study strategies to achieve their goal, so to engage them in activities that will encourage them to evaluate themselves and be able to explicitly explain what they know and do not know. Self-assessment provides potential for novice learners to improve their understanding.

**Self-questioning**

Self-questioning is an aid to metacognition, it can describe a process through which learners ask and answer questions while reading. This process enables them to understand and digest text information and become independent learners who can actively engage with organized thinking through goal-directed learning. Researchers in the metacognition field suggest that engaging in self-questioning is part of attributes of skilled reading (Williamson, 1996). Self-questioning is a type of skills that relates to self-management strategy that can be utilized to change behaviors, guide instruction, gain understanding, complete tasks, and many more (Joseph, Alber-Morgan, Cullen, & Rouse, 2016). In many cases, difficulties in comprehension are always related to failure of readers to actively participate in the reading process. To increase novice programmers’ understanding of the important information from their reading sources as well to increase their motivation, self-questioning is one of the best approaches in practice. (M. Rum et al., 2014), for example, had designed reflective activities in his computer-assisted learning system to encourage self-questioning about learning experiences leading to self-directed learning. The more students used self-questioning in various kinds of situations, the more likely they were to develop self-questioning habits; this ultimately became an automatically-triggered skill used subconsciously as required in a given situation.

**Self-directed learning**

Self-directed learning (SDL) has received increased attention, particularly in the higher education context. (Malcolm Shepherd Knowles, 1975) advanced the most common foundation definition of SDL as a process in which individuals take the initiative without external help to manage their own learning, identify learning needs, formulate learning goals, implement appropriate learning strategies, apply required resources, and evaluate learning outcomes. The widely-known benefits of SDL are, traits of people who take initiative and responsibility in learning and learn more and better than those who do not (Malcolm Shepherd Knowles, 1975). (Gureckis & Markant, 2012) pointed out that SDL helps learners optimize their educational experience, allowing them to focus on useful information. The active nature of SDL also helps learners in the process of encoding information.
and retaining it over time. Other benefits of SDL have been espoused by a number of theorists in adult education (Bollnus, 2003; Malcolm S Knowles, Holton III, & Swanson, 2014). SDL is related to relationships between learners, self-directed learning, and the impact of technology (e-learning, internet and broadband) on self-directed learning (Hiemstra, 1994). In this research work, as suggested by (Mishra, Fahnoe, & Henriksen, 2013), the fact that the learners take initiative and responsibility for learning process, allowed them to manage, select and assess their own learning activities can be touted as self-directed learning.

Graphic organizers

Contemporary technologies appear to provide many opportunities for addressing the learning needs of novice programmers with respect to difficulty in learning computer programming concepts. (Fouh, Akbar, & Shaffer, 2012) suggest that the practice of using graphic representation of an algorithm provides a logical abstract for aiding learners in developing the logical thinking required in computer science courses. Figure 1 is an example of graphical representation of a data structure overview as an example of fundamental learning in computer programming.

Figure 1. Graphical representation for data structure overview (Chansilp & Oliver, 2006)

Graphic representations have long been used to solve variety of problems and textual understanding. This technique can help learners analyze text by representing the structure and flow of textual information. Flowcharts, Venn diagrams, concept maps, and tree diagrams are examples of graphic organizers that can be used to better understand textual information. Network tree, cycles, fishbone maps and series and continua/scales are some identified graphic organizers useful in text-reading activity. A graphic organizer definition and common types of graphic organizers are presented in Figure 2.

Figure 2. Graphic Organizer definition and common graphic organizer types (Otto & Everett, 2013)

Relationships between text elements and conceptual text structures can be illustrated using graphic organizers (Stull & Mayer, 2007). A graphic organizer is a text adjunct that is “a meaningful diagram formed from statements and words that graphically-connected is called a visuospatial information arrangement” (Horton,
Lovitt, & Bergerud, 1990). Concept maps, knowledge maps, Venn diagrams, causal diagrams, and matrices are just a few of the common graphic organizers found in textbooks (Stull & Mayer, 2007) and utilized in the classroom. When attempting to comprehend and learn from expository text, readers must be able to employ both surface-level and deep strategies. Surface-level strategies are those that help readers understand and remember fact-level text-based and macro-level information (Kintsch, 2005). Deep processing strategies help readers select important information, link new information to prior knowledge, and ultimately construct a situation or mental model (Kintsch, 2005). Graphic organizers help readers to achieve efficient learning and facilitate the process of knowledge construction through explicit connection with text concepts (Horton et al., 1990). Graphic organizers are particularly applicable to metacognitive reading strategy instruction and self-regulated learning because they can be utilized before, during, and after reading to monitor metacognitive and assess their learning.

**Experimental studies**

A system called Metacognitive Support Environment was developed to measure the effectiveness of proposed application support features. The motivation behind the development of this system was desire of novice programmers for an environment supporting learning computer programming metacognitively. The system support features as discussed in previous sections, i.e., scaffolding (support tailored on student’s needs), self-questioning (encouraging the process of asking and answering while learning), self-assessment (the process of having the learners critically reflect upon and record the progress of their own learning), graphic organizer (used to organize information during learning) and self-directed learning (learner take the initiative without external help to manage their own learning) and reflective prompt (provides directive help for learner). The system is comprised of the following five main activities.

*Pre-Stage (reflective stage, i.e., what learner’s know and don’t know) - The objective is to make students reflect both on what they know and do not know; this activity takes place before the process of learning begins. It presents appropriate conditions to make a student realize the importance of using suitable and possible strategies, providing reference resources as well as the degree of focus required for success in solving a problem. The main objective of this stage is for students to be triggered by their own reflections to monitor their knowledge. It focuses on their past experience in solving problems as well as the performance level indicated by low, average, or high. Students will be able to compare their estimation and judgment of their own actually-understood knowledge through exposure to activities such as knowledge monitoring and performance comparison as well as analysis of knowledge monitoring. These activities stimulate and encourage self-directed learning in which students are forced to take initiative and responsibility for learning.*

*Familiarization stage* (assess learner’s understanding) –the stage in which students both assess their understanding and develop a strategy for problem-solving. During this stage, students are presented with a list of possible strategies as most appropriate for solving the problem; alternatively they themselves can compose new strategies. The primary objective of this stage is to make them reflect on strategies that may help them solve the problem. This activity emphasizes metacognitive strategies that relate to the process of solving the problem. Students will be exposed to activities that allow them to compare their own knowledge judgment with their actual knowledge and help them select the strategies provided for solving the problem. One type of metacognitive support element provided during this stage is the reflective prompt used to provide directive help for learner in the learning process. Figure 3 presents an example of reflective prompts and graphical organizer used to parse postfix expression into an expression tree. In this learning example, learners were asked to build an expression tree based on the given postfix expression. The operations buttons were provided for the learner to choose to structure the expression in the graphical form.

*Production stage* (learner’s self-assessment and problem comprehension) is the stage in which students perform self-assessment and deal with problem comprehension. In this stage, students are required to solve the given problem by presenting an answer. The main objective in this stage is to reflect an understanding with respect to the concept as well as demonstrating confidence in correctly solving the problem. This activity relates to the student’s performance self-assessment. Activities taking place during this stage concentrate on translating the given problem into pseudocode and monitoring the application of the planned strategies. The activities are comprised of problem-solving, checking answers, and quizzes. Students are thereby exposed to self-assessment activity where they must evaluate their own work.

*Evaluation stage* (learner’s experience of past problem solving) - the stage in which students evaluate their own experience in past problem solving. This stage involves only the activity of checking the solution provided by the lecturer, used as a comparison for studying the student solution. Graphical representation would be used to
convey a learner’s metacognition information in the form of reflectometers to trigger knowledge monitoring accuracy and time management (see Figure 3).

**Figure 3. Reflective prompts and Graphical organizer used to convert postfix expression into an expression tree**

**Post-task stage** (learner’s to reflect on the problem solved) – The main objective of this stage is to have the students reflect on the problem solved by providing an opportunity to review their most recent experiences as well as to explore things that happened during the activity of solving the problem. Students should be able to identify the “cause of the mistakes” that occur during the problem-solving activity, the time spent, and the resources used. As suggested by (Quintana, Zhang, & Krajcik, 2005) in their framework for scaffolding tools, a progress bar is provided during this stage to enable students to see the their most recent activity timelines.

**Figure 4. Evaluation dashboard screenshot**

**Figure 5. Post-task screenshot**
Figure 4 is a screenshot of evaluation dashboard, and Figure 5 is a screenshot of post-task activity; these are elements of the proposed system.

A discussion of development of the proposed support system can be found in (M. Rum et al., 2014; S. N. M. Rum & Ismail, 2014). The primary target of the proposed system is to scaffold the knowledge monitoring skill of a programmer, i.e., to provide the ability and opportunity to assess one’s knowledge, or by extension, one’s understanding. We believe that promoting awareness of knowledge monitoring accuracy at the novice level is the first step in fostering metacognitive skills that in turn help in making a good selection of strategies and facilities for better allocation of cognitive resources.

Participants

An experimental study was conducted with two different groups of subjects: a control group and an experimental group each associated with different conditions. A four-week training course, Introductory Computer Programming, was given to each group. Throughout the training sessions, only the experimental group was exposed to and interacted with the proposed support system, while respondents in the control group were given only ordinary class sessions without interaction with the proposed support tool. Sixty-six of 100 target respondents (first year Computer Science undergraduate students) were invited via email and agreed to take part in this study. They were randomly assigned into two groups namely experimental and control group. From the total of 66 participants, 30 agreed to take part in the experimental group and 36 in the control group. The distribution of respondents is shown in Table 1.

<table>
<thead>
<tr>
<th>Group</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>30 (Male = 12, Female = 18)</td>
<td>45</td>
</tr>
<tr>
<td>Control</td>
<td>36 (Male = 15, Female = 21)</td>
<td>55</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>100</td>
</tr>
</tbody>
</table>

Method with research design

The experimental design had a pre-test and post-test similar to the underlying idea for evaluating the (Tobias, Everson, & Laitusis, 1999) model of metacognition. These tests were designed to measure basic knowledge concepts related to programming and knowledge monitoring. A pre-test was given to both groups before the training session and a post-test given for both groups after the training session. The tests (both pre- and post-test) were divided into two parts. In the first part, for both tests, 10 minutes was given to the students to complete the tests; they were asked to estimate their knowledge in terms of whether or not they would be able to solve the problem given. For each problem, they were asked to answer “yes” or “no” to the question: “Do you think you can translate the given problem into pseudocode?” In the second part, again for both pre- and post-test, they were required to translate the same problems into pseudocode. All participants (pre- and post-test) were given 1 hour to complete these tests. The post-test was designed to be more difficult than the pre-test, given that the subjects should have gained more practice in solving problems during their training. Both tests (pre-test and post-test) had five questions in total and were devised with the assistance of a computer programming instructor. Pseudocode is an artificial and informal language that helps programmers develop algorithms. Pseudocode is “text-based” (algorithmic) and written in English as a detailed yet readable description of what a computer program or algorithm must do, expressed in a formally-styled natural language rather than in a programming language. Pseudocode makes creating programs easier and allows programmers to concentrate on the logic of the problem to be solved (the algorithm) without having to know programming language syntax in detail. It depicts the entire algorithm’s logic so that it requires only rote implementation to translate it line by line into executable source code.

Procedure and measurement instruments

The scoring system was as follows: 2.0 points for a correct answer (the logic of the written pseudocode is totally correct), 1.0 point for a partially correct answer (the logic of the written pseudocode is partially correct), 0.5 points for an incomplete answer (the logic of the written pseudocode is incomplete) and 0 for an incorrect answer (the logic of the written pseudocode is totally incorrect). Scoring was performed by computer programming experts, i.e., faculty lecturers, each with more than 5 years of experience in teaching computer
programming. For measuring the knowledge monitoring assessment of learner (designated KMA). This instrument have been used by many researchers (Moran, 2012; Sangin, Molinari, Nüssli, & Dillenbourg, 2011) to evaluate student knowledge monitoring ability across domains. We adapted the empirically-validated instrument developed by (Tobias & Everson, 1996) after changing it slightly by giving a little additional flexibility to the possibility that a student predicted that they would partially solve the problem or that they partially understood it. Table 2 presents the knowledge monitoring assessment results with score values of a, b, c, d, e, f, g, h, i. The KMA exhibited nine possible scores reflecting the relationship between a student’s prediction of their knowledge and her actual performance. A score of 1 was given for the situations a, e and i, while a score of -1 was given for situations c and g, and a score of -0.5 was given for situations b, d, f, and h. The mean of knowledge monitoring scores over all the problems solved represents the current KMA state of the Student.

Table 2. KMA Score Condition

<table>
<thead>
<tr>
<th>Actual performance</th>
<th>Unable to solve it</th>
<th>Able to solve it partially</th>
<th>Able to solve it</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides incorrect answer</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>Provides partial correct answer</td>
<td>d</td>
<td>e</td>
<td>f</td>
</tr>
<tr>
<td>Provides correct answer</td>
<td>g</td>
<td>h</td>
<td>i</td>
</tr>
</tbody>
</table>

Results

Wilcoxon tests were performed to investigate the repeated measures of difference between the groups. In this study, if the research hypothesis is true, it is expected that positive rank is higher in experimental group and similar numbers for both positive and negative rank in control group. An examination of the findings shown in Table 3 reveals that there was a significant difference between the pre-test and post-test scores of students in the experimental group (Z = -2.292, p = .000 < .001). The sum of negative ranks for the experimental group students’ academic achievement scores was 7.50, while the sum of positive ranks was 58.5. There were 20 respondents show an increase in terms of performance improvement in learning Introductory Computer Programming before and after training session, only 2 of respondents from experimental show a drop on performance and 8 respondents were not affected with the four-week training with the aid of proposed system throughout the training session. Given the sum of ranks for the difference scores, the result shows that there was a significant improvement in post-test scores as compared to pre-test scores obtained by the experimental group.

Table 3. Wilcoxon test results for pre- and post-test scores changes of experimental group

<table>
<thead>
<tr>
<th>Ranks</th>
<th>N</th>
<th>Mean rank</th>
<th>Sum of ranks</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PostTest - PreTestExp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Ranks</td>
<td>2a</td>
<td>7.50</td>
<td>7.50</td>
<td>-2.292</td>
<td>.022</td>
</tr>
<tr>
<td>Positive Ranks</td>
<td>20b</td>
<td>5.85</td>
<td>58.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ties</td>
<td>8c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *PostTest < PreTestExp; †PostTest > PreTestExp; ‡PostTest = PreTestExp.

The results in Table 4 show that there was no difference between the pre-test and post-test scores of the control group, where the sum of the negative ranks of the control group was 2.5 and the sum of the positive ranks was 1 with Z = -1.089, p = .0276 at the 5% level. There were 30 respondents in this group did affected with the given training session, 4 respondents did not show a drop in performance before and after the training session, only 2 respondents show positive on the improvement of performance in learning Introductory Computer Programming. Given the sum of ranks for the difference scores of the control group, the result shows that there was no difference of performance changes in control group for the four weeks training.

Table 4. Wilcoxon test for pre- and post-test score changes of control group

<table>
<thead>
<tr>
<th>Ranks</th>
<th>N</th>
<th>Mean rank</th>
<th>Sum of ranks</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PostTest - PreTestExp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Ranks</td>
<td>4a</td>
<td>2.50</td>
<td>5.00</td>
<td>-1.089</td>
<td>.276</td>
</tr>
<tr>
<td>Positive Ranks</td>
<td>2b</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ties</td>
<td>30c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

Research into the challenges of teaching and learning programming to novices has long received attention to many researchers especially when dealing with conceptually rich domains that require metacognitive perspective
as a vehicle to stimulate the learning process. In this study, we proposed a metacognitive educational environment for learning Introductory Computer Programming for novice programmers. We have restricted our definition of novices to tertiary level students who are learning programming for the first time. Parse problem solving in the form of pseudocode before translating into actual programming language is a common method in computer programming pedagogy. Respondents that participated in this study were novice programmers randomly selected and assigned into two groups; control group and experimental group. Both groups were received four weeks training on Introductory Computer Programming, however, only the experimental group had been exposed to and interacted with the proposed support system throughout the training session. In this study, we have developed the hypothesis that the more student interact with the proposed system the more reliable the proposed support features of the system. The KMA was used as the instrumental measurement of student performance in learning Introductory Computer Programming. In Table 3, the result of experimental group shows that 66.7% (20) of respondents show an improvement in post-test after received four-weeks training and exposed to and interacted with the proposed support system throughout the training session, while 26.7% (8) of the respondent did not show any improvement, meaning that there is no differences of performance changes between pre-test and post-test and only 6.6% (2) respondents show an inclination on the performance in learning Introductory Computer Programming subject. In Table 4, the result show that there was no differences of performance gained by the majority of respondents in control group (84%) (30) Before and after the training session, only 5% of the respondents show an increase in performance after received the training session, while 11% of the respondents show an inclination on the learning performance. From the result, it can be concluded there were positive effects in terms of performance gained in learning computer programming with the aid and interaction of the proposed support tool during the training session. This was might be due to the effect of metacognition support elements provided in the support tools. The finding of this research work support the hypothesis and consistent with the studies done by other researchers (Azevedo, Moos, Johnson, & Chauncey, 2010; Bernard & Bachu, 2015; Feng & Chen, 2014; Ismail, Ngah, & Umar, 2010). The result demonstrates that the proposed metacognitive strategies of computer-assisted learning support (metacognitive scaffolding, reflective prompts, self-assessment, self-questioning, self-directed learning and graphic organizer) helps learners in improving their learning performance in computer programming, especially in knowledge monitoring and problem-solving.

**Conclusion and future study**

In the present study, we have discussed characteristics that should be considered for developing computer-assisted learning tools in a metacognitive learning environment for supporting teaching and learning of computer programming. The present study support the hypothesis, that with the aid metacognitive strategies, student would be able to achieve their goal in learning computer programming, perhaps, in order to have better understanding the learning success is to view and measure all other factors such as learner motivation state, cognitive strategies and etc. This is the possibility of future investigation of how other factors contribute to learning success in computer programming subject. It is also important for more research to be conducted in this area with stronger controls in order to show whether or not metacognitive accuracy can be improved further. For the improvement of this research work, it is also suggested for future work to get more respondents to be participated as well as to increase the duration of training session (e.g., one semester) to see the effectiveness of the proposed system. We summarize the benefits of metacognitive strategies as discussed in this research work in Table 5.

<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metacognitive scaffolding</td>
<td>Provides coaching, guides and advice</td>
<td>Strengthens the transferability of digital learning by helping to improve novice learner adaptability to a particular learning situation</td>
</tr>
<tr>
<td>Reflective prompts</td>
<td>Provides directive helps for learner during learning process</td>
<td>Encourage learners to reflect on techniques and strategies that they employ in learning process</td>
</tr>
<tr>
<td>Self-questioning</td>
<td>Learners ask and answer while learning</td>
<td>To aid learner’s in comprehension monitoring strategy; it is very useful in assisting novices to become self-directed and independent.</td>
</tr>
<tr>
<td>Self-directed learning</td>
<td>Learners take initiative and responsibility in learning process</td>
<td>Learner more willing and motivated to learn because of the ability to manage, select and assess his/her own learning activities.</td>
</tr>
<tr>
<td>Graphic organizer</td>
<td>Graphical representation used to solve a variety of problems.</td>
<td>The complexity of a computer programming subject can be reduced by helping learners in developing the logical thinking required by computer science courses.</td>
</tr>
</tbody>
</table>
References


Lai, G. (2008). Examining the effects of selected computer-based scaffolds on preservice teachers’ levels of reflection as evidenced in their online journal writing (Unpublished doctoral dissertation). Georgia State University, Atlanta, GA.


Investigation of Continuous Assessment of Correctness in Introductory Programming

Deller James Ferreira¹*, Hebert Coelho da Silva¹, Tatiane F. N. Melo² and Ana Paula Ambrósio¹

¹Instituto de Informática, Universidade Federal Goiás, Câmpus Samambaia, Goiânia, Goiás, Brazil // ² Instituto de Matemática e Estatística, Universidade Federal Goiás, Câmpus Samambaia, Goiânia, Goiás, Brazil // dellerg@inf.ufg.br // hebert@inf.ufg.br // tmelo@mat.ufg.br // apaule@inf.ufg.br

*Corresponding author

(Submitted March 9, 2016; Revised August 4, 2016; Accepted October 25, 2016)

ABSTRACT
Teachers usually expect that any form of a continuous assessment (CA) should contribute significantly to the student learning process in introductory programming courses. To foster teachers to go beyond the current practices of a CA applied when to programming, from 2011 to 2014, we investigated the use of the Boca Online Contest Administrator (BOCA) system, an online judge used in programming marathons, and the CA of program correctness in the cohorts of an introductory programming course. Empirical results show that there is no significant difference between student’s performance when comparing the cohorts that used a CA and did not use a CA, and when comparing the cohorts that used and did not used the BOCA system but used a CA. An in-depth analysis revealed the potential and limitations of the use of a CA and the BOCA system, unveiling the need for the adoption of assessment practices and environments that build cumulative knowledge through multiple means of assessment, allowing profitable interactions among students, and between students and their teacher regarding the students’ solutions.

Keywords
BOCA system, Continuous assessment, Introductory programming

Introduction

Learning how to program is notoriously complex and difficult. Computer educators share the assumption that their students find programming difficult to learn (Jenkins, 2002; Bornat, 2011; Dijkstra, 1982). In addition, programming demands several skills that are intertwined, and the novice programmer needs to deal simultaneously with multiple processes.

Solutions to computer programming problems should be modular, efficient, clear, validated, and accurate. In addition to being a creative activity, students have to learn the syntax and semantics of programming language elements, and how to combine them into meaningful programs (Muller, Haberman & Averbuck, 2004; Robins, Rountree & Rountree, 2003; Utting et al., 2013). To succeed, students must develop the ability to deal with computational problems regarding the many creative and cognitive aspects of programming.

In addition to difficulties in programming, student performance is frequently below the teacher’s expectations (McCraken et al., 2001). For example, Jerinić et al. (2014) assessed the grades of students taking introductory programming at different institutions and in various countries, and concluded that the students in introductory programming courses do not know how to program at the expected skill level.

Many researchers have considered different perspectives to the teaching and learning of programming (Koulouri, Lauria & Macredie, 2014; Soloway & Ehrlich, 1984; Sphorer & Soloway, 1986; Tang, 2009; Denadhi, 2009). However, there is still no revolutionary pedagogy to teaching programming, and no consensus on what is the best way to learn it. Meanwhile, most teachers still apply a traditional teaching method for introductory programming courses that consists of lectures and programming exercises.

There is empirical evidence that programming exercises can have an even more important role than simply applying the theory taught during lectures. Exercises can be seen as teaching artifacts that complement lectures by teaching the same content but in an exploratory manner.

In an attempt to make the learning of programming more effective, feedback should be provided to the students during their programming exercises because they learn more effectively when they know what is expected from them. Students look forward to feedback that enables them to improve as learners. During programming activities, feedback is often given by means of assessment tasks (McCraken et al., 2001; Earl & Uscher, 2012).
An assessment is any act of interpreting or acting upon information regarding a student’s performance, which is collected by a variety of means or practices (Nicol & Macfarlane-Dick, 2006; Messick, 1989). An assessment can significantly influence the effectiveness of student learning (Hattie & Timperley, 2007; Vihavainen, Airaksinen & Watson, 2014). Educators should consider applying assessments as an integral and important component of the teaching and learning process.

For the purpose of boosting student learning, a CA is seen as an ongoing process arising out of the interactions between teaching and learning. An assessment involves both the teacher’s and the student’s use of information, and provides evidence of student performance. The main aspect of the assessment process is to enhance an evaluation by means of a CA that can imply the student’s level of progress (Fisher & Frey, 2007). A CA can help a student become more aware of any gaps that exist in their learning process, and can motivate them to narrow these gaps before taking an exam.

Feedback can offer students an experiential base for reflection. For the purposes of this work, we consider reflection as a mental process that incorporates critical thought regarding an experience. A student’s ability to reflect on the materials they have produced, in order to form reasoned judgments, is central to a deeper level of learning. An assessment should provide insight to the students to enhance their learning processes by allowing them to identify their strengths and weaknesses. By means of such reflection, the students can practice and demonstrate self-knowledge to create a new and different course of action.

However, assessing the students and providing appropriate feedback to them in terms of their programming are not simple tasks. Some educators examine and grade their students’ assignments manually, whereas others prefer automatic tools to ease the efforts required for the assessment. Automatic programming assessment systems have recently become a significant method in assisting educators to automatically assess and grade their students’ programming exercises. Among these tools is the Boca Online Contest Administrator (BOCA) system (de Campos & Ferreira, 2004).

Many researchers have highlighted the evidence indicating that an automatic programming assessment can provide data to increase student performance. These data are directly related to the criteria chosen to assess a computer program. Given a particular problem to be solved by the students, there are a number of criteria that can be used to assess the computer program used to solve the problem.

An automatic programming assessment system can assess whether a program is correct. In addition to correctness, some assessment tools are used to analyze the program efficiency, coding style, and the existence of inline documentation. However, there are other important criteria used to judge the level of programming. Because programming is creative and cognitively complex, and because there are constraints imposed by the programming language used, it is important for the student to receive data regarding the act of programming, and not only data judging the program itself. The use of automatic tools focuses mainly on the correctness of the solutions and neglects other important properties such as the design and clarity of the program.

For example, the students need to know whether the chosen data structure is appropriate, whether the chosen control flow is the best, how to structure the code used to solve a problem into smaller units (which can potentially be re-used), and what algorithm or programming patterns are going to be applied, combined, and adapted. For both the students and teachers, the variety of different correct solutions can be a challenge.

In addition, an international study on introductory computer science courses conducted by Carter et al. (2010) showed that most teachers assess the programming exercises submitted by their students merely for their correctness, either manually or using an automatic assessment tool.

Thus, considering that certain works highlight the need to embrace criteria other than the correctness during a CA of the programming level, and being aware that providing only feedback regarding the correctness of a program is a current practice adopted by teachers, further research that aims to provide a deeper understanding about what an assessment of correctness is, and what this type of feedback can obtain from the students regarding their achievements, is essential. Capturing and highlighting the educational benefits and limitations of an assessment of the correctness in programming are required to better enlighten the teachers.

In this work, we investigate the continuous assessment of correctness in introductory programming. The results indicate that, despite the correctness of their code being an essential aspect of a programmer’s work, an assessment of the correctness provides important but insufficient data for effective feedback in promoting student reflections and in helping teachers develop facilitation strategies.
Related works

Teaching can be considered a closed-loop feedback process, where the teachers require an achievement measure of their students’ learning to be aware of the weaknesses and strengths of their students, and to overcome such weaknesses (Ashour et al., 2014). Clearly, an assessment applied during the learning process is valuable, and any evaluative works conducted prior to the teaching are deemed worthwhile.

According to Serra-Toro, Traver, and Amengual (2014) despite continuous assessments being commonly used in higher education, some of their limitations are often ignored. Continuous assessments can influence the ways in which students approach their studies, and therefore contribute indirectly, but effectively, to the quality of their learning, but only when carefully designed.

Nevertheless, the relationship between assessment practices and the overall quality of teaching and learning is still undermined. There is often a lack of assessment requirements and assessment criteria assuring the effectiveness (Brown & Hattie, 2012).

According to Hattie and Timperley (2007), the type of feedback can cause a greater or lesser degree of effectiveness. Feedback in programming can even fail, as reported through previous research (Soloway & Ehrlich, 1984; Scott et al., 2015). Scott et al. (2015) observed that continuous feedback is not always efficacious, depending on certain factors such as the amount, nature, and timing of the feedback.

The continuous characteristic of the assessment does not guarantee student progress. Successful continuous assessments must display evidence regarding the level of learning at a sufficiently early time period, in such a way that the teaching and learning can be modified, causing greater progress toward the intended goals and targets. Progress involves being able to perform faster, more accurately, and more easily; knowing more; and achieving a deeper programming capability (Scriven, 1967).

An adequate measurement of a student’s learning performance should provide opportunities for improvement, along with guidance on the redesign of the course objectives. Assessments applied to improve the degree of learning must be used to diagnose an individual’s learning difficulties and suggest remediation procedures.

Ihantola, Karavirta, and Seppälä (2011) made a systematic literature review regarding systems for an automatic assessment of programming assignments. They indicated that an automatic assessment has certain limitations. One of the limitations of current automatic assessment tools, such as the BOCA system, is the limited quality of the automatic feedback generated.

The use of automatic testing allows the students to be accurate and obtain the maximum score, but does not guarantee that several of the program requirements are checked. Automatic tools cannot examine whether a variable name is meaningful, if the program is well designed, or even if the program uses a specific data structure. Automatic tools focus mainly on checking the solutions, and neglect other important aspects. In addition, the final product (program) tends to be emphasized over the process (programming).

A CA, whether automatic or not, can contribute to the student learning process only if the data used for the feedback, or used to generate the feedback, come from a set of relevant assessment criteria, including not only the criteria for a program assessment, but also the criteria for judging the act of programming itself.

Implementing a CA properly requires the teachers to reconsider their role, their students’ roles, and their interactions with their students. Interactive techniques for a description and identification of the causes of programming errors need to be applied more frequently. Teachers need to watch the behavior of their students while programming, and listen closely to their conversations. At times, they may need to ask questions during conversations to clarify details regarding what the students are doing and what they are discovering, but otherwise, they should not interfere with the students. Teachers need to ask purposeful questions that enable the students to reflect upon, clarify, and explain their thinking and actions. In addition, it is profitable to immerse the students in collaborative learning, where student dialogue is valued.

An investigation into the application of a CA of correctness, either using an automatic programming assessment system or not, is a relevant contribution. To delimit the influence of providing feedback regarding the correctness of a program for the learning of novice students, revealing the potentialities and limitations of the program is a necessary issue to motivate teachers into incorporating interactive assessment procedures and feedback strategies to assure the quality of learning how to program.
Research questions

To reveal the limitations and potentialities of a CA of correctness in introductory programming, we conducted a comparative analysis by considering the cohorts when the BOCA automatic programming assessment system and a manual CA of the program correctness were and were not applied, with the objective of enlightening and encouraging teachers to go beyond adopting such instructional practices. The research questions applied are as follows:

RQ1. Do students in cohorts, where the teacher applies continuous assessment and the programs are checked for correctness, attain better educational achievements?

RQ2. Do students in cohorts, where the teacher applies the use of BOCA for program assessment, reach better educational achievements?

Method

A comparative study was conducted with the goal to investigate the use of the BOCA system and a CA of correctness in an introductory programming course. To answer the research questions, an empirical study was conducted using statistical techniques to analyze the results obtained from an assessment of the students’ performance.

The BOCA system is an automatic program correction system commonly used in Brazil for checking programs using predefined inputs and outputs. The BOCA system receives the students’ submissions and uses the given inputs to verify whether a program’s output is identical to the expected output for the given input. The BOCA system serves as a repository of solutions submitted by the students, where feedback regarding the correctness is provided.

Quantitative measures for analyzing data have been addressed, and empirical methods have been employed. An empirical method reveals observational data, making it possible to make inferences considering the effects of the use of the BOCA system and the application of a CA, and to conduct a comparative analysis between cohorts in which the BOCA system and a CA were and were not applied.

The statistical techniques used to compare the different educational situations relevant to the above research questions are the Quantile-Quantile (QQPlot) graph and the Mann-Whitney test. The empirical data measurements used were the final exam grades and ongoing assessment testing grades.

QQPlot generates a graph used to compare the characteristics of two populations, and allows a very intuitive visualization to check whether there are differences between the grades of the cohorts. If there are no significant differences between the grades, the dots are close to the diagonal line, inclined at 45º, passing through the origin. The Shapiro-Wilk test showed the non-normality of our data at the 5% level, which lead us to opt for the Mann-Whitney ($U$) test. The Mann-Whitney ($U$) test is an alternative to the t-test that does not require the data to follow a normal distribution. Unlike the t-test, which compares the mean values of two groups, the Mann-Whitney ($U$) test compares their medians. It examines the differences between two independent groups on a continuous scale.

With the aim of eliciting the reasons for the students’ level of performance and statistical results, we also conducted an in-depth analysis aimed at detecting how the feedback generated by assessing the correctness of the programs triggered the reflections of the teacher and students, and resulted in new chains of actions aiming to attain better learning outcomes. We elaborated on two questionnaires, one focusing on the students, and the other focusing on the teacher, with the intent to generate qualitative data for an in-depth analysis.

Participants and procedures

This study was carried out during introductory programming courses, which were offered to the students during the first semester. None of the students had previous experience with programming when they started the course. Each semester, there are 40 students entering the computer science program. For the introductory programing courses, the students are always subdivided into two groups. This division allows the teacher to provide greater attention to the students. In some cases, the cohorts exceeded the expected number of 20 students because some of the students were repeating the course.
The programming language, which was used between 2011 and 2014, was the C programming language, and the course structure covered elementary data types, simple and structured commands, arrays, and matrices. The choice of C as the programming language was dictated by the pedagogical requirements of the course. Because we strove to maintain the variations in the variables to a minimum, using a single programming language allowed for a more consistent environment for comparing the assessment approaches. However, we do not believe that the choice of a different programming language would have had a major implication regarding the results because the feedback given during the assessment was basically whether the solution is correct or not. Any differences would have been due to the characteristics of the language and not to the assessment approach.

The study, conducted from 2011 to 2014, comprised 136 undergraduate students, and involved six cohorts and one teacher. The notation used here is YYYY/S, where the first four digits correspond to the year (YYYY), followed by the semester (S). For example, 2012/2 stands for cohort 2012 and the second semester.

Table 1 presents the analyzed cohorts, showing the number of students in each cohort (N) and their median final grade. In addition, the table shows which cohort used a continuous assessment (CA), and in which evaluation tasks the BOCA system was used. Depending on the cohort, the BOCA system was applied for CA tasks, exams, or exercises. Some cohorts had a CA but did not use the BOCA system for checking their correctness. The CA tasks were programming exercises that the students solved using a pencil and paper in class. The grades for the cohorts with no CA were composed of three exams taken during the semester. The grades for cohorts with a CA were made up of three exams and several small quizzes taken each week. Each student conducted all of the activities. For all of the cohorts, the students received only feedback regarding the correctness of their program.

In classes where the BOCA system was used, there were some particularities in the number and variety of activities undertaken with the BOCA system. For cohort 2011/2, all nine continuous assessment tasks were carried out using the BOCA system; six, using an automatic response; and three, with the automatic response option disabled. During all other activities (exercise lists, classroom exercises, etc.) the BOCA system was not applied. Cohort 2012/1 had an average of six continuous assessment tasks that did not use the BOCA system, and three exams that did. All other activities were carried out without the BOCA system. For cohorts 2012/2, 2013/2, 2014/1, and 2014/2, BOCA was used during all activities. Extra class activities were left on an online BOCA server, allowing the students to do them from home. In addition, at the beginning of the semester, a list of 40 exercises was left on the online BOCA server for students to do during the semester.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>N</th>
<th>Median</th>
<th>CA</th>
<th>BOCA</th>
<th>CA</th>
<th>Exams</th>
<th>Exercises</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011/2</td>
<td>25</td>
<td>2.44</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2012/1</td>
<td>29</td>
<td>5.52</td>
<td>Yes</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2012/2</td>
<td>19</td>
<td>3.65</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2013/2</td>
<td>23</td>
<td>1.92</td>
<td>No</td>
<td></td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2014/1</td>
<td>18</td>
<td>2.58</td>
<td>No</td>
<td></td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2014/2</td>
<td>22</td>
<td>2.82</td>
<td>No</td>
<td></td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

For all of the cohorts, the programming activities assigned to the students were extracted from the Brazilian Informatics Olympiad (Anido & Menderico, 2007). These exercises involved real-world problems whose solutions are computer programs. Within the problem definition, strict rules for the input and output data formats were established. Such strict rules are important for two reasons.

First, to specify a problem’s solution (computer program), the students must abstract the data and requirements that are often not fully explicit in the problem definition. This makes the students apply their abstraction capability. Second, the student’s programs must strictly obey the input and output rules defined in the exercise, allowing the BOCA system to evaluate the programs properly.

Learning outcome measurement of correctness

The learning outcome assessment of correctness addressed the following measurement criteria from 2011 to 2014:

- The grades varied from 0 to 10.0, and students needed to score at least 5.0 to pass.
- When a CA was applied, the final grade was given by the following formula: (average grade in ongoing tests + average grade in final exams)/2.
When a CA was not applied, the final grade was given by the following formula: average grade of the final exams.

Considering the application of a CA when using or not using the BOCA system, a program that correctly processed the datasets was considered totally correct and earned the maximum grade; otherwise, the program earned a zero. The maximum grade for an ongoing test is 10.0, i.e., the sum of all programs is 10.0.

Regarding the final exams, each exam contained four programs. The maximum grade for each program was 2.5. A program that correctly processed the datasets for multiple inputs earned a grade of 1.5. A program that reacted properly to an erroneous input earned an additional 0.5. A program that was compiled and run without errors earned another 0.5.

Questionnaires

The questionnaires were elaborated upon to extract qualitative data originating from the teacher’s and the students’ user experiences, new courses of action, and reflections on the content, knowledge, skills, abilities, and attitudes. The data collected from the questionnaires inform us the general level of satisfaction with the usage of the BOCA system, for example, whether the teacher found the BOCA feedback useful for improving their teaching, and if the BOCA feedback allowed immediate students or the teacher to make adjustments in response to student difficulties. Both questionnaires are described as follows.

Student questionnaire

1. Did the use of BOCA entice you to develop new forms of studying?
2. Did the use of BOCA encourage you to assess your skill levels?
3. Did the use of BOCA help you to detect your weaknesses?
4. Did the use of BOCA trigger any reflections on your performance?
5. How do you feel when using BOCA?
6. Did the use of BOCA motivate you to seek new content?
7. Did the use of BOCA motivate you to solve more exercises?
8. Did you like using BOCA?
9. Do you think you learned better with BOCA?
10. Did the use of BOCA generate group discussions?
11. Did you discover new things when using BOCA?
12. Do you think that BOCA meets your particular method of learning?
13. Did you overcome obstacles using BOCA?
14. Did the use of BOCA help in your development?
15. Did the use of BOCA lead you to try harder in subsequent tasks?
16. Did the use of BOCA bring about actions to improve your performance?
17. Was the use of BOCA a pleasant experience?
18. What was the overall balance from the use of BOCA?
19. Did the use of BOCA lead you to seek help from your teacher or classmates?
20. Did the use of BOCA involve you to engage in collaborative work with your classmates?

Teacher questionnaire

1. Did the use of BOCA lead you to develop new forms of feedback for students?
2. Did the use of BOCA help you to detect student weaknesses?
3. Did the use of BOCA trigger reflections on the performance of the students?
4. Did the use of BOCA motivate you to deliver new content?
5. Did the use of BOCA motivate to you deliver more exercises?
6. Do you think that you taught better when using BOCA?
7. Did the use of BOCA motivate you to hold discussions with the students?
8. Did you discover new things when you used BOCA?
9. Do you feel that BOCA meets your method of teaching?
10. Did you overcome any educational obstacles using BOCA?
11. Did the use of BOCA help you in your teaching?
12. Did the use of BOCA lead you to try harder in subsequent teaching activities?
13. Did the use of BOCA result in actions to improve your teaching?
14. Was the use of BOCA a positive experience?
15. What was the overall balance from using BOCA?
16. Are the data generated from BOCA sufficient to formulate effective teacher feedback to the students?

**Empirical results**

To answer the research questions of this work, we conducted a study in which we interpolated qualitative and quantitative techniques to analyze the data obtained from the students’ grades, as well as data that came from the students and teacher questionnaires.

**Statistical results**

Considering those cohorts that used and did not use the BOCA system, and those cohorts that used and did not use a CA, the median distribution of the final grades are visualized in Figure 1. As shown in Figure 1, there was a slight difference in grades, indicating better results in classes that used a CA and in classes where the Boca system was not used. However, it is necessary to check whether these differences are statistically significant.

![Figure 1. Visualized median of the final grades when applying/not applying a CA and the BOCA system](image)

The grade distribution was analyzed by means of the Mann-Whitney \((U)\) test. The Mann-Whitney \((U)\) test was conducted in order to answer the first research question (RQ1), considering null hypothesis \(H_0\): there is no significant difference between the median of the final grades of the cohorts that used a CA (2011/2, 2012/1, and 2012/2) and those that did not (2013/2, 2014/1, and 2014/2). The pertinent test is \(H_0: \mu_X = \mu_Y \times H_1: \mu_X \neq \mu_Y\). Here, \(\mu_X\) and \(\mu_Y\) are the medians of cohorts X and Y, respectively, where X is the group of students that used a CA, i.e., cohorts 2011/2, 2012/1, and 2012/2. In addition, Y is the group of students that did not use a CA, i.e., cohorts 2013/2, 2014/1, and 2014/2. When comparing the grades of the cohorts, the result of the statistic test \(W\)
was 3331.5 and the $p$-value was 0.22, indicating a level of 5%, which means that, for the majority of the cohorts studied, no significant difference was found between the scores of those who used a CA and those that did not (Figure 2).

Additionally, the Mann-Whitney ($U$) test was conducted in order to answer the second research question (RQ2), with null hypothesis $H_0$: there is no significant difference between the median of the final grades of the cohorts that used the BOCA system (2011/2 and 2012/2) and those that did not (2012/1). The settings for these assumptions are the same as previously described.

The Mann-Whitney ($U$) test showed that there is no significant difference in the distribution of grades among the cohorts' instances, indicating that the feedback from a CA regarding the correctness of a program given to the students when using or not using the BOCA system did not help the students to increase their grades.

Figure 3 shows QQPlots for the previous configuration. Observed at a level of 5%, we can conclude that there is no significant difference between the median scores of X (where BOCA was used) and Y (where BOCA was not used). The result of the statistic test $W$ is 551.5 and the $p$-value is 0.33.

**Figure 2.** QQPlot comparing the average scores obtained for cohorts where a CA was used, and for cohorts where a CA was not used

Figures 2 and 3 show QQPlot graphs for situations related to the research questions. QQPlot graphs are used to compare the characteristics between two populations. QQPlot graphs contain dots representing the quantiles of each sample. If two samples come from the same population, then the dots should be along the 45° diagonal line starting at the origin. The samples can be compared based on their distribution, checking whether the points plotted on the graph are distributed near the 45° diagonal line. When the points plotted on the graph are parallel to the diagonal line, the two distributions are similar. We adopted the conventional $a = 5\%$ probability to determine a statistical significance.

An interpretation of Figures 2 and 3 corroborates the idea that the use of a CA when applying or not applying the BOCA system was not more effective, but when a CA was applied, the results were slightly better. In both situations, the points representing the grades are distributed near the diagonal line. We must mention here that in both situations where the BOCA system was used, the teacher was applying a CA. However, there were differences regarding the intensity of the feedback and the method used in correcting the programs. The
assessment of correctness was automatic, and the student feedback regarding the program correctness was given more frequently.

![QQPlot comparing the grades of cohorts that used BOCA and the cohorts that did not. In all cohorts a CA was performed](image)

*Figure 3. QQPlot comparing the grades of cohorts that used BOCA and the cohorts that did not. In all cohorts a CA was performed.*

It is worth mentioning that, as shown in Table 1, the median was 5.52 for cohort 2012/1, which is much higher than those of the remaining cohorts. Considering that for cohort 2012/1, the Boca system was not used for either a CA or the exercises, we can state that this is evidence reinforcing the results.

We expected that the continuous feedback regarding the level of correctness was not going to improve the students’ grades much, given its limitations. However, we did expect that automatic and more frequent feedback of correctness could boost the students’ performance, even to a small degree. To determine the reasons for the apparent inconsistency in the results, we delivered a questionnaire to the students and teacher focusing on whether the BOCA system could trigger the students’ and teacher’s reflections and actions, and if so, to what extent.

**In-depth analysis**

For a qualitative analysis, an in-depth analysis was conducted on the data collected from the questionnaires with the aim of enlightening the statistical results. The in-depth analysis covered the students’ and the teacher’s reflections, actions, and user experience.

Regarding the students’ reflections, 82.6% of the students affirmed that the use of the BOCA system encouraged them to assess their skills, 73.9% said that the use of BOCA helped them detect their weaknesses, 67% answered that they learned better with BOCA, 42% discovered new things when using BOCA, 63.3% responded that BOCA meets their method of learning, 73.9% overcame obstacles using BOCA, 76.5% replied that the use of BOCA helped in their development, and 64.2% thought that their performance was improved. These results indicate that the students’ perceptions regarding their improvement in knowledge were positive. In addition, all of the answers were positive, indicating that the use of BOCA led them to reflect on their failures.
Concerning the students’ actions, 69.8% developed new forms of study using BOCA, and 70.3% answered that the use of BOCA led them to try harder in subsequent tasks. These facts indicate that BOCA lead the students to embrace a reactive attitude with the objective of overcoming their difficulties. This positive outcome, which is related to the students’ actions along with the fact that BOCA also boosted their reflections, assures us that BOCA had a positive impact in student learning, and has potential to trigger student reflections and actions. Thus, we can conclude that the cause for the lower grades of students in cohorts in which the BOCA system was applied was not the absence of student reflections or corrective actions. These results corroborate the importance of providing feedback regarding the correctness to the students, showing that it has the potential to boost the students in making forward steps toward a deeper level of learning.

A possible explanation for the decreased performance of the students when using BOCA was the lack of collaborative and interactive practices. Only 30.9% of the students stated that the use of BOCA generated group discussions, 34.1% replied that the use of BOCA led them to seek help from their teacher or classmates, and 32.9% responded that the use of BOCA induced them to be involved in collaborative work with their classmates. Based on this, we have evidence showing that students had a bad social experience. They took individual actions to overcome obstacles, but were not motivated to work in groups when using the BOCA system. The use of the system could have facilitated a more individualized study, in detriment to collaborative learning and interactions with the teacher.

In relation to the students’ experience during the use of the BOCA system, 63.2% felt enthusiastic when using BOCA, 50.8% affirmed that the use of BOCA motivated them to seek new content, 63.8% responded that the use of BOCA motivated them to solve more exercises, 59.2% liked using BOCA, 60.9% had a pleasant experience while using BOCA, and 50.8% considered the overall balance when using BOCA to be a good experience. These results indicate a positive affect concerning the use of BOCA. Most of the students enjoyed themselves while programming when using BOCA. In addition, they were motivated to use the system. We can therefore discard a lack of motivation as a possible cause for the lower performance of those students that used the BOCA system.

Students included additional comments highlighting some of the shortcomings related to the BOCA system’s inability to show what actually is wrong in their learning when a program is incorrect. These observations reinforce our previous argumentation concerning the limitations of simply providing feedback regarding the correctness of a program, in which only the program is evaluated, but not the learning processes involved during the act of programming. In addition, a lack of student interactions and collaborations could have been the cause for the students’ poor performance when using BOCA. These results are in line with what was previously addressed in this work. We deemed the attention given to student interactions and collaborations as essential to learning how to program. Some student excerpts are as follows:

“BOCA works well for programming Olympiads, but not for learning, where knowing what really went wrong helps a lot.”

“BOCA does not show how we fail.”

“BOCA does not evaluate the learning, it only evaluates the goal.”

With respect to the teacher reflections, the teacher responded that the use of BOCA triggered reflections on the performance of the students, but the teacher did not discover new things when using BOCA.

Regarding the teacher actions, the teacher answered that the use of BOCA lead him to develop a few new form of feedback for his students, and that it helped him to detect student weaknesses. However, he was not able to overcome educational obstacles using BOCA, and the use of BOCA slightly helped him in teaching, led him to try slightly harder in subsequent teaching activities, and resulted in few actions improving his teaching. In addition, the data generated from BOCA did not provide sufficient information to formulate effective teacher feedback to the students.

Concerning the teacher’s experience, the teacher replied that the use of BOCA did not motivate him to deliver new content, but did motivate him to provide more exercises. In addition, he considered that he taught better with BOCA on only a few occasions, and that the use of BOCA motivated him to have greater discussions with the students, but also on only a few occasions. He considered that BOCA meets his method of teaching and was a positive experience, and he considered the overall balance from the use of BOCA to be good.
In general, the teacher’s responses indicate that the use of BOCA neither led him to make an effort to try and teach better, nor provided sufficient information to understand the students’ limitations. However, he considered the use of BOCA to be a good experience and a practice that meets his method of teaching. This is evidence that the teacher is accommodated and does not seek to become a better teacher. The results from the teacher questionnaire lines up with the idea that teachers must be warned regarding the limitations of only assessing the correctness of a program, and that they should be motivated to seek new interactive methods for assessing the students in order to unravel the educational problems that lead the students to build an incorrect program.

Conclusions

The objective of this work was to motivate teachers to develop more interactive teaching and learning practices, establish links between empirical results and key issues regarding programming feedback, and going slightly further, address the limitations and potentialities of using an assessment of program correctness as a vehicle for improving student learning.

Empirical research conducted from 2011 to 2014 in an introductory programming course gave us insight into common CA practices, indicating that there is a need for better evaluation processes for determining student performance and ways for dealing with relevant assessment information. This research highlighted the need for resources and practices that go further than simply applying ongoing tests, and providing only feedback for a program’s correctness when using or not using the BOCA system.

Empirical results show that there is no significant difference between the students’ performance when comparing the cohorts that used and did not use the BOCA system and a CA, indicating that a CA using BOCA with only automatic feedback for correctness, or applying a CA where the teacher simply checks for correctness, does not significantly improve the teaching and learning processes of computer programming. The Mann–Whitney test did not indicate a difference at a 5% level between the cohorts that used and did not use BOCA, and between the cohorts that used and did not use a CA. Although this difference is visible in our dataset, the result is not statistically significant.

Automatic program correction systems are widely used in programming marathons or Olympiads worldwide. This is due mainly to the large number of submissions to be assessed within a short period of time. Certainly, the use of systems such as BOCA assist in giving feedback more quickly, and in undergraduate programming disciplines, quick feedback to students is extremely important. However, the quality of the feedback should also be taken into account, and BOCA feedback is necessary for triggering student reflections and corrective actions, but is limited and insufficient to imply a significant improvement in students. Thus, we believe that the use of BOCA is important, but should be accompanied by interactive teaching strategies and the teacher’s facilitation of collaboration among the students.

Another issue of discussion is whether the use of a CA regarding program correctness helps teachers and students to correct any mistakes early in the teaching/learning processes through more knowledge evaluation checkpoints. A large number of assessments and quick feedback on student performance triggered a slight amelioration in student performance. However, this was proven to be an inefficient practice regarding the teaching and the students’ understanding of the reasons that lead to an incorrect program.

We suggest that diversified teacher/student interactions using BOCA can aid in student achievement, as long as automatic feedback from BOCA is used sparingly. Given that automatic feedback basically indicates whether a solution is right or wrong, a student who submits a program with an error, even if the program contains only small mistakes, will not find any support helping to fix the student’s program. Thus, teachers must interact with their students to be aware of which problems cause programming mistakes, and provide quick feedback with better quality than BOCA can offer.

In conclusion, the use of BOCA and the application of a CA for verifying the correctness of a program correspond to a minimally informative practice. The use of BOCA and the application of a CA regarding only program correctness are a necessary but insufficient assessment method. Noteworthy efforts are required to improve the use of a CA in practice.

The aim of this work was to foster teachers to go beyond simply checking for the correctness of their students’ computational solutions. The adoption of assessment practices and environments that build a cumulative knowledge through multiple means of assessment is required, allowing valuable interactions among students,
and between the students and their teacher, regarding the students’ solutions. Teaching reflection needs to be more interactive, investigative, and integrated into the students’ individual methods of learning. There is a necessity for the teacher to develop interactive and collaborative strategies to reveal to the teachers and students data related to chains of cognitive breakdowns that lead to programming errors.

**References**


A Study of Supplementing Conventional Business Education with Digital Games

Abida Ellahi¹*, Bilal Zaka² and Fahd Sultan³

¹Fatima Jinnah Women University, Rawalpindi, Pakistan // ²Virtual Education, COMSATS University, Islamabad, Pakistan // ³COMSATS University, Islamabad, Pakistan // abia.ell@gmail.com // bilal@vcomsats.edu.pk // sfsultan@gmail.com

*Corresponding author

(Submitted November 11, 2016; Revised February 27, 2017; Accepted March 23, 2017)

ABSTRACT

This paper documents how the adoption of digital games by academia reshapes the current worldview by bringing the potential answers for all learning issues. The central objective of this study is to investigate the extent to which digital games can impact learning effectiveness, and to what extent these games can be used as supplementary elements for existing pedagogical methods prevailing in Pakistan. The study used experimental research method to investigate the effects of digital games on learning outcomes of students in business education in Pakistan. The statistical analysis was done using Partial Least Square method with the help of SmartPLS Software. All results confirmed that students in higher education sector are ready to accept and adopt new technologies that can better facilitate their learning process. They are ready to create and share their knowledge in a collaborative manner by using technological platforms.

Keywords

Experimental research, Digital games, Business education, Satisfaction, Performance, Instructors’ support

Introduction

“If we teach today’s students, as we taught yesterday’s, we rob them of tomorrow” - John Dewey (1915)

Technology in education has been on upsurge during recent years. The modern advancement in educational technologies has given different opportunities to educational establishments for embracing new types of learning. The improvement of effective learning in a computer or technology mediated learning environment is an exciting challenge. The major challenge postulated by the technology in education is that students who are born in this computerized and versatile digital age participate in learning from an exceptionally assorted point of view than their forerunners (Pollara, 2011). These types of students were called by Prensky (2001) as “digital natives, who are at low compatibility with faculty called as ‘digital immigrants’.” The low compatibility between the teacher and students brought by new educational technologies, stresses a need for revising the existing pedagogical strategies so that learning outcomes of the students can be enhanced.

Pakistan has lagged the world in adopting the educational technologies as the use of innovative tools in education are at their infancy. For this country, where there are not many learning advancements, studies exploring the potentials of technologies in education are crucial. Hence, this study has tried to investigate the extent to which new learning technologies can influence learning adequacy in Pakistan and to what degree these advances can be utilized as supplementary components for the existing teaching strategies. For this purpose, digital games have been tested in business education studies in a university setting. The study was planned to investigate the effects of digital games on learning satisfaction and perceived learning performance of business graduates with moderating role of instructor support.

Literature review

Digital games in education

Majority academic and professional institutes have identified the need for new forms of learning, beyond traditional learning and training methods. Among these new forms of learning or teaching methods, Digital Game-based Learning (DGBL) has been gaining popularity and attention (Rondon, Sassi & Andrade, 2013). This Digital Game-based Learning has been considered as a different instructional method to increase learning which cannot be addressed by the traditional lecture-discussion instruction. The digital games are not only interesting for children but for teachers and researchers as well due to their rich features representing a real environment (Squire, 2002). Researchers and developers are trying to explore new ways to exploit the interactive
potential of digital games in the learning environments within the boundaries of psychological and philosophical beliefs about learning process (Hannafin, 1992). However, the effectiveness of digital game based learning still needs clarification and validation for generalizability.

Games containing educational objectives are thought to be helpful, more interesting, and, thus, more effective in the learning of academic subjects (Popescu, Romero & Usart, 2013). For this reason, academic world has been trying to provide personal learning environments (PLE) by introducing new technologies like games (Gåsland, 2011). For example, “reference schools such as Harvard Business School have meant to make a progress, from the evolution of paper based study cases into simulations and interactive case studies where the learners could play a realistic situation, to learning by doing” (Srikant, Garvin & Cullen, 2010). Many researchers have been using different research designs and theoretical frameworks for explaining and measuring digital game based learning. Although these studies have provided enough evidence about the effectiveness of digital games in learning, there do not exist sufficient facts about such games in higher education (Orzcelik, Cagiltay & Ozcelik, 2013).

Despite many premises about the benefits of digital game-based learning (e.g., Steinkuehler, 2008), little empirical evidence exists about the effectiveness of such games in higher education as the existing studies do not adequately address the relationships between integrating digital game-based learning in a university curriculum and students’ learning. Although, success of serious games has been shown in many recent studies (e.g., Hwang & Wu, 2012; Wouters et al., 2013), the real prospective of these games in education is “still far to be fulfilled, concerning higher-order learning goals” (Connolly et al., 2012) and “there is a growing need for educational technology research in this field” (Bellotti, Bottino, Fernández-Manjón & Nadolski, 2014).

Activity theory

Activity theory is a psychological and multidisciplinary theoretical framework that has its roots in the Russian psychology. The original framework of this theory was developed by the Russian psychologist Aleksei Leontiev (Leontiev, 1978). Engeström (1987) further illustrated six components of this triangular structure of an activity system namely Subject, Object, Tools, Community, Division of Labour, and Rules. The assumption of this theory is based on interrelationship of the subject (the learner), the object (the goal which leads to the outcome), and the tools (both physical and conceptual) used to mediate between them. Activity theory conceives that activity system (e.g., family, a religious organization, a political movement, a course of study, a school, a research laboratory) is the basic unit of analysis of behaviour in individuals or collective. It argues that the relationship between objects in the environment and people are intervened by culture and its rules, the community, and by labour and its roles and development. The central part of activity theory is the hierarchical framework of activity that is composed of: activity, actions and operations, characterized by objective, goals and conditions, respectively (Leont’ev, 1981).

Activity theory developed decades ago has been widely used for explaining human computer interaction. The activity theory offers a comprehensive framework when information technology is used in higher education for students and teachers (Hashim & Jones, 2007). Activity theory has been applied to games (Dobson et al., 2005; Squire, 2002), and has been studied in the context of players’ learning. Oliver and Pelletier (2004) using activity theory, designed a framework that facilitates the tracking of learning without disturbing the natural flow of game play. The game playing process in digital game based learning environment enhances the learning by allowing players to acquire learning experiences in games, encouraging interactions between learners and the game system as well as situating learners in complex learning environments (Pannese & Carlesi, 2007). The game playing processes are interesting to play (Huang & Johnson, 2009), thus promote meaningful learning.

Learning satisfaction

Tsai, Yu and Hsiao (2012) investigated the factors that enhance learning effectiveness in digital game based learning environment. They found that learning motivation, learning ability and playing skills are direct determinant of learning effectiveness in digital game based learning environment. Kelle, Klemke and Specht (2013) in an experimental study, found strong support of design patterns for learning games on the learning outcome and user experience. Yien, Hung, Hwang and Lin (2011) in a quasi-experimental non-equivalent-control group design found positive effect of game playing on learning achievements and learning attitudes of students in nutrition education. Liao and Wang (2011) investigated students’ usage of business simulation games and its effect on their motivation, satisfaction and intention to use. They found satisfaction as a significant
determinant of continuous usage intentions, hence, satisfaction must be monitored in game-based learning contexts. Therefore, it can be hypothesized that:

H1. Use of digital games for learning significantly affects perceived learning satisfaction of the students.

Perceived learning performance

Ariffin, Oxley and Sulaiman (2014) defined learner performance as “increase of knowledge and capability of learner as a result of learning activity.” They evaluated the effectiveness of using digital games for learning in higher education and found that a student’s background affects his motivation to learn which in turn affects his performance. Zafar, Mueen, Awedh and Balubaid (2014) explored the relationship between using computer games and students’ academic learning performance. They used a game based learning with native language hint and found that students using games for learning performed better than those who did not. They also highlighted that students’ enjoyment while playing games significantly affect their learning performance. Hence, it can be hypothesized that:


Learner interest

Li (2010) discussed that “game motivation” encompasses four concepts i.e., interest, anxiety, probability of success, and challenge. The game playing process in digital game based learning environment enhances the learning by allowing players to acquire learning experiences in games, encouraging interactions between learners and the game system as well as situating learners in complex learning environments (Pannese & Carlesi, 2007). The game playing processes are interesting to play (Huang & Johnson, 2009), thus, promote meaningful learning. Farrell (2005) demonstrated a learning environment, including simulations, in an international business course and argued that simulations having participative, inductive, interactive, reflective, and exploratory characteristics along with traditional pedagogical methods stimulate a student’s interest and engagement to achieve the expected learning purposes. Therefore, interest has been taken as a mediator between digital game play and learning outcomes in this study. Hence, it has been hypothesized that:

H 3(a): Learner’s interest mediates the effect of use of digital games for learning on learning satisfaction of the students.

H 3(a): Learner’s interest mediates the effect of use of digital games for learning on perceived learning performance of the students.

Instructor support

The chore of incorporating games into an instructive setting is a challenging one, which requires instructors to arrange many organizational resources. “Beyond the practicalities of ensuring that game sessions run reliably from an administrative perspective, the teacher also needs to be able to guide and support students’ gaming experiences during activities. Being a game tutor for students entails several responsibilities for the teacher, and given the variation in individual students’ proficiencies and interests, this task can be rather difficult……game-based learning processes are demanding on teachers, requiring them to take on many different roles, each of which requires a specific skillset” (Marklund & Taylor, 2014).

Taylor (2015) addressed the issue of instructor role in game based learning environment. She said that in game based learning literature, focus has been made solely on game characteristics and teacher’s role has been overlooked. However, instructor-led serious gaming requires many roles such as facilitator, debriefing role, and coach or in-game organiser, player/participant, off game enabler, leader, expert, and technical support. Needs for the teacher’s roles should be effectively met for the smooth running of serious gaming (Taylor, 2015). McDaniel and Telep (2009) found that “students may be enticed by the thought of playing a video game as one of their assignments, but unless instructors provide a clear and critical introduction to the assignment and a debriefing period, students might, in the end, deem the gameplay experience as “filler” or even as an instructor’s attempt to pander to their likes and desires. “ In this study, instructor’s support in digital games based learning environment has been tested as a moderating variable. Hence, it has been hypothesized that:
The proposed research model is shown in Figure 1.

![Proposed research model](image)

**Figure 1. Proposed research model**

**Research design and methodology**

To assess the value of the proposed method, an experimental designed study was conducted. This experimental study investigated the effectiveness of digital game based learning in business education. As this specific study is concerned with the evaluation of attitudes of students toward digital game, hence, it would not have been feasible to set up multiple control groups (Connolly, Stansfield & Hainey, 2011). Therefore, the experimental design used for the study was pre-test game post-test.

**Research population and sample**

The population participating in this study was students of BBA (Bachelor in Business Administration) Program. The sample of the experiment consisted of the students of the two classes who had joined fifth semester of their BBA Honours degree program and were undertaking Operations Management course. Hence, the study was set in the context of a core business education curriculum subject. To be included in the study, students should have successfully completed the introductory classes of Operations Management, especially the topic of project management and had to be proficient in computer usage. Prior to their enrolment, all students who participated in the study were informed of the purpose and procedures, after which all showed their willingness. In this experimental set up, students from the Operations Management course were randomly assigned to these groups. A total of 47 students for the experimental group were recruited from three classes of the same course, while 40 students were included in the control group. The student profiles were similar to each other. They were at the same state of educational, gender and age range. As the university selected was a women university, hence, all the participants were females. Subjects were provided with continuous assistance and support for queries throughout the gaming session. Subjects were allowed to share and discuss with each other their progress and scores. In short, game was played in a socially collaborative environment. One class was assigned to the experimental group (N = 47), while the other class was assigned to the control group. The two classes were taught by the same instructor, so that different teaching methodology does not affect the results. The students in experimental design learned the project management topic using the game, while the other group was taught in a conventional way.

**Procedure**

For this experimental design, there were two classes of students who were assigned to the pre-test Game post-test experimental design. Prior to their actual interaction with the game, the subjects were given brief oral instructions on its use by the researcher. Throughout the interventions, the researcher also observed students’ interaction with the game and also provided procedural help to the students. The gaming session was conducted in a classroom where each participant was seated in front of a laptop computer (Figure 2). The researcher arranged
for Wi-Fi Internet as the game was to be played online. Before starting the game session, a pre-test was taken from both groups. This test contained particular questions about the game topic. After the gaming session, a post test was taken from the experimental group. In a subsequent teaching hour, the subjects were administered a questionnaire. Subjects filled the questionnaires anonymously, in their classrooms, in the presence of the researcher. Some of the students who were absent in that class but attempted the gaming session returned the questionnaire via email. The time needed for completion of the questionnaire were approximately 20-30 minutes.

**Instrument**

Both, the subject test and questionnaire, were used to assess their performance and opinions. Since the validation of the instructional effect of using games was the main objective of the study, pilot test was managed. Subject experts were involved in carrying out pilot testing to ensure validity and reliability of criterion measurement of questionnaire. All variables were rated on 5 point Likert scale. This study adopted the measures used to operationalize the constructs from the previous literature, with minor rephrasing to tailor these measures to the game based context.

**Material: My Sust House: The Game**

The game selected for this study was My Sust House which is a collection of exciting interactive games related to environment, building and the town which can be used for different subject areas like project management and operations management. The game dealt with an eco-friendly house building project. In the words of Architecture and Design Scotland (2006) “The environment game explores ways to create a more sustainable environment. The building game challenges the children to build a sustainable house. Player receives a printable certificate with their score.” This game helps to apply critical thinking by making informed decisions. These games are equally suitable for science, social studies and technological studies. These games can be accessed at http://www.mysusthouse.org/.

![Figure 2. Students playing games in the classroom](image)

**Data analysis and results**

In order to analyse the data of this experimental study, partial least square method was employed using SmartPLS software. The SPSS software was also used to analyse pre and post test scores of the assessment test. An independent sample t-test was run before starting the game session in the experimental group. The pre-test was taken as a knowledge test about the project management of a house building. The results did not show a significant difference \((M = 4.6, SD = 2.59, t(85) = 0.37, p = .71)\), between the control group and experimental group as shown in Table 1. The independent sample t-test was further run to find the difference between pre-test and post-test knowledge test between the two groups. The statistics show that subjects who played game for learning scored higher \((M = 6.71, SD = 1.55, t(85) = 8.29, p = .000, \text{significant})\) than those who did not play game \((M = 3.22, SD = 1.76, t(85) = 8.29, p = .000, \text{significant})\).
Table 1. Results of pre and post test score

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>df</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control group</td>
<td>40</td>
<td>4.65</td>
<td>2.60</td>
<td>.37</td>
<td>85</td>
<td>.710</td>
</tr>
<tr>
<td>Experimental Group</td>
<td>47</td>
<td>4.45</td>
<td>2.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control Group</td>
<td>40</td>
<td>3.23</td>
<td>1.76</td>
<td>-8.29</td>
<td>85</td>
<td>.000</td>
</tr>
<tr>
<td>Experimental Group</td>
<td>47</td>
<td>6.17</td>
<td>1.55</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PLS path model estimation and evaluation

To test hypothesized relationships among variables in game based learning context, path analysis was done by using Partial Least Square (PLS) analysis with the help of SmartPLS 3.0 trial version software. The conceptual model of the current study comprised 5 latent variables with 30 observed indicators. Due to the sample size of 47 in this experimental study, the covariance based modelling would have been unfeasible. The PLS model by using the SmartPLS 3.0, data was analysed in two stages: first, the measurement model was assessed for adequacy and secondly, the structural model was weighed.

Measurement model

The statistics for measurement model are shown in Table 2. The values consist of AVE, Composite Reliability (CR) and Cronbach Alpha values. Composite Reliability and Cronbach Alpha represent the same reliability values. CR has been considered as an alternative to Cronbach Alpha in PLS modelling. Garson (2016) described that Average Variance Extracted (AVE) may be used as a test of both convergent and divergent validity. AVE reflects the average communality for each latent factor in a reflective model. The values given in the Table 2 show that composite reliabilities CR of all five latent constructs were all above 0.80, meeting the minimum criteria. The values of AVE, as a measure of convergent and discriminant validity, are all meeting the requirements of minimum 0.5.

Table 2. Measurement model statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>AVE</th>
<th>CR</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Game play</td>
<td></td>
<td></td>
<td>0.495</td>
<td>0.906</td>
<td>0.885</td>
</tr>
<tr>
<td>2 Learner interest</td>
<td>0.390</td>
<td>0.377</td>
<td>0.562</td>
<td>0.834</td>
<td>0.730</td>
</tr>
<tr>
<td>3 Learning satisfaction</td>
<td>0.699</td>
<td>0.671</td>
<td>0.706</td>
<td>0.878</td>
<td>0.791</td>
</tr>
<tr>
<td>4 Learning performance</td>
<td>0.693</td>
<td>0.664</td>
<td>0.550</td>
<td>0.859</td>
<td>0.793</td>
</tr>
<tr>
<td>5 Instructor support</td>
<td></td>
<td></td>
<td>0.456</td>
<td>0.868</td>
<td>0.827</td>
</tr>
</tbody>
</table>

Structural path estimation

The structural model was tested by estimating and testing the significance of the structural path coefficients (direct effects) and then their indirect effects of the latent variables through mediating and interaction terms. The mediation test was conducted by observing the significance level of the indirect paths that emerged from the independent to the dependent variables, using the bootstrapping procedures incorporated in SmartPLS.

The results of mediation analysis for both learning performance and learning satisfaction as dependent variables are shown in Table 3. In this test, level of significance of the indirect paths both in the absence of the intervening variable (total effects, denoted C paths) and in its presence (direct effects, denoted $C^0$ paths) were examined. The values in Table 3 shows direct, indirect and total effects of independent, intervening and dependent variables.

The statistical values for mediation analysis in Table 3 show that the independent variable, game play has significant effect on both learning performance ($\beta = 0.71, t = 7.69, p = .000$) and satisfaction($\beta = 0.77, t = 9.38, p = .000$). This independent variable also has a significant effect ($\beta = 0.63, t = 8.59, p = .000$) on the mediating variable which is Learner Interest. Hence, assumptions for path “c” and “a” had been met. However, when it came to path “b” that is the effect of mediating variable, Learner Interest on dependent variables, it did not come out significant i.e., for Learning Performance ($\beta = 0.11, t = 1.34, p = .18$) and for Learning Satisfaction ($\beta = 0.16, t = 1.40, p = .16$). In this way, the third assumption about the mediation effect turned out as insignificant. Hence, in this examination, the initial assumptions of a complete mediation analysis could not be met. The Learner Interest might be a good variable in the model still, but it is not mediating as theorized. Hence, the hypotheses H1. Use of digital games for learning significantly affects perceived learning satisfaction of the students and H
2: Use of digital games for learning significantly affects perceived learning performance of the students have been accepted.

**Table 3.** Hypotheses testing of learning satisfaction and perceived learning performance

<table>
<thead>
<tr>
<th></th>
<th>Direct effect</th>
<th>Indirect effect</th>
<th>Total effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>( t )</td>
<td>( p )</td>
</tr>
<tr>
<td>1</td>
<td>Game Play ( \rightarrow ) Learning performance</td>
<td>.641</td>
<td>5.363</td>
</tr>
<tr>
<td>2</td>
<td>Game Play ( \rightarrow ) Learning satisfaction</td>
<td>.668</td>
<td>6.587</td>
</tr>
<tr>
<td>3</td>
<td>Game Play ( \rightarrow ) Learner Interest</td>
<td>.625</td>
<td>8.596</td>
</tr>
<tr>
<td>4</td>
<td>Learner Interest ( \rightarrow ) Learning performance</td>
<td>.110</td>
<td>1.344</td>
</tr>
<tr>
<td>5</td>
<td>Learner Interest ( \rightarrow ) Learning satisfaction</td>
<td>.158</td>
<td>1.404</td>
</tr>
<tr>
<td>6</td>
<td>Instructor Support ( \rightarrow ) Learning performance</td>
<td>.174</td>
<td>1.813</td>
</tr>
<tr>
<td>7</td>
<td>Instructor Support ( \rightarrow ) Learning satisfaction</td>
<td>.108</td>
<td>1.017</td>
</tr>
<tr>
<td>8</td>
<td>IS’LP ( \rightarrow ) Learning performance</td>
<td>-.073</td>
<td>0.991</td>
</tr>
<tr>
<td>9</td>
<td>IS’LP ( \rightarrow ) Learning satisfaction</td>
<td>.152</td>
<td>1.683</td>
</tr>
</tbody>
</table>

*Note.* Game Play \( \rightarrow \) Learning Satisfaction \( R^2 = 0.69 \), Game Play \( \rightarrow \) Learning Performance \( R^2 = 0.693 \), Game Play \( \rightarrow \) Interest \( R^2 = 0.390 \).

**Figure 3.** Path analysis of variables
The path analysis values are shown in the Figure 3 while Table 3 contains the values of beta, t value and significance level for the effects of variables on each other. In the first step, it was assumed that game play for learning has a positive effect on learning satisfaction of students. This hypothesis was supported by the values indicated in the Table 3. The beta values \((\beta = 0.67, t = 6.59, p = .000)\) confirm that Game Play significantly affects the Learning Satisfaction of students. The second assumption was about the positive effect of Game Play on students’ perceived Learning Performance. The resulting \((\beta = 0.64, t = 5.36, p = .000)\) values again confirmed the positive and significant effect of Game Play on students’ perceived Learning Performance. The effect of game play on creating learner’s interest was also assumed positive. The values of \((\beta = 0.63, t = 8.59, p = .000)\) again confirmed this assumption as well. The direct effect of Learner’s Interest on Learning Satisfaction and Learning Performance did not come significant i.e., \((\beta = 0.16, t = 1.40, p = .16)\) and \((\beta = 0.11, t = 1.34, p = .18)\) respectively. Hence, assumption about mediation effect of Learner’s Interest could not be scrutinized or confirmed. Hence, the hypotheses H3(a): Learner’s interest mediates the effect of use of digital games for learning on learning satisfaction and H3(b): Learner’s interest mediates the effect of use of digital games for learning management system for learning on learning performance have been rejected.

**Moderation analysis**

Moderation analysis was also done in this study, as this study used one moderator. The inclusion of moderator in a model, alters the strength of relation between two variables. The effect of Instructor Support was examined on both dependent variables that are Learning Satisfaction and Learning Performance. The statistics shown in Table 3 indicate that the direct effect of Instructor Support on Learning Performance was \((\beta = 0.17, t = 1.81, p = .007)\) and on Learning Satisfaction was \((\beta = 0.108, t = 1.017, p = .310)\) which were clearly not significant. As the main effect was not significant, hence, the interaction term did not turn out to be significant on both Learning Performance and Learning Satisfaction (i.e., for Learning Performance, \((\beta = -0.073, t = 0.99, p = .32)\) and for Learning Satisfaction \((\beta = -0.152, t = 1.68, p = .093)\). Hence, the moderation effect did not turn out to be significant in this analysis. Hence, hypotheses H4(a): Perceived instructor support moderates the relationship between use of digital games for learning and learning satisfaction and H4(b): Perceived instructor support moderates the relationship between use of digital games for learning and perceived learning performance have not been supported.

**Discussion of findings**

The study examined the effect of employing games in pedagogy on students’ satisfaction and their perceived learning performance via a mediating role of learner’s interest and a moderating role of instructor’s support. In order to answer the research questions, a pre and post-test assessment was made from the students’ control and experimental group. The score differences indicated that subjects who were introduced to games, had seen a significant change in their knowledge after playing the game and subjects in control group who did not play the game did not show much difference in pre- and post-test. After playing the game, students in the experimental group were asked to express their opinions on a Likert scale questionnaire, so that theoretical relationships could be examined. The statistical results show that use of digital games enhances students’ Learning Satisfaction and Learning Performance. This indicates that students who play games to support their education, are more satisfied with their learning outcomes and with this instruction method. They believe that this tool and technology would bring significant change in their learning performance.

These findings suggest that blending an interactive game-based approach with traditional classroom delivery increases learning effectiveness of students in higher education. These findings are consistent with many previous studies for example a study by Kanthan and Senger (2011) confirmed that use of digital games for education improved academic performance of the students, increased their satisfaction and engagement and decreased their stress as well as “foster an improved, facilitated, fun, nonthreatening, extended student learning environment.”

The current study also examined the mediating role of learner’s interest in this game play environment. It was assumed that using game for learning actually creates interest among learners about the technology as well as learning and this interest creates satisfaction and their perceived learning performance. When this assumption was tested statistically, it was found that game play undoubtedly increases learner’s interest like Ting (2010) confirmed in his study that using a game in teaching, increases student’s interest in learning and Zafar, Mueen, Awedh and Balubaid (2014) in their study also concluded that students using games reported more enjoyment and better understanding of subject.
On contrary to expectation, the effect of this learner’s interest on students’ satisfaction and perceived learning performance could not bring significant results. Hence, Learner’s Interest could not be proven as a mediator in this study. This result was consistent with one of the study of Seo and Baek (2010) who established an indirect effect of games fantasy on achievements in learning via interest, intrinsic motivation, and storytelling as mediator variable. He found that although interest is the “strongest factor of fantasy in educational games, but it did not show a significant effect on academic achievements.”

Although the mediating effect did not bring significant results, however, Learner’s Interest still appeared as a significant variable in the context of game base learning. These unexpected results may have been due to a low sample size, or distinctive features of the study and game. Irrespective of the theoretical role of this variable, the findings clearly suggest that variance in use of game can bring Learner’s Interest. This, in turn, prove that interest is a valuable part of game based learning environment, although, it was not a mediator as theorized in this research. One finding can be drawn that game play has a strong direct relationship or effect on Learning Satisfaction and perceived performance, rather than having indirect effect. Hence, it was found that Learner’s Interest should be taken as a separate outcome of game play along with satisfaction and performance rather than having a mediator, like Meesuk and Srisawasdi (2014) studied a direct effect of game play on increased students’ flow, enjoyment, learning, satisfaction and motivation. Future researches could use this study’s results to investigate the dependency of effect of game play on learning satisfaction and performance is on learner’s interest or not.

This study also tested moderating effect of Instructor Support to strengthen the relationship of game play and learning outcomes among the students. In this research, it was assumed that when students receive instructor support in a game based learning project, it strengthens the effect of the game play on their learning outcomes, like satisfaction and performance. Contrary to the expectations and literature support, the instructor support in this game based environment could not turn out to be a significant variable or having a moderating effect. González-Cruz et al. (2003) examined the different levels of instruction support i.e., detailed, intermediate and minimal. They found that although instructional support is necessary in using computer simulations, however, the students must have some freedom while using this simulation, rather than having teacher’s “still reviews and comments on their work afterward.” Due to this, they suggested to use the “intermediate level of instruction, where both freedom and structure are offered.” In the light of these previous findings, it can be inferred that instructor support could not be proven as a significant moderating variable does not indicate a zero effect or an exclusion of the variable. In fact, it shows that the level of instructor support needs to be carefully studied. It is not instructor support as a whole, in fact it is the level of instructor support that matters in bringing learning outcomes in game based sessions as Podleschny (2012) observed “the teacher role in relation to the gameplay was indeed not fixed, but constantly re-configured by the actions and interactions of the network”

By applying the activity theory, subjects found in this study were players of the game and the artefact used to mediate the activity was digital game. The prime motive to be engaged in this activity was better learning effectiveness and the outcomes of this activity fulfilled the objectives i.e., positive learning outcomes. Regarding the community, game was played in a physical environment, where traditional way of interaction took place among the members. The rules in this activity system were rules set by games programmers. The primary division of labour within this game based environment was students as “active learner” and teacher as “facilitator.” The theoretical lens of activity theory provided a context to understand and relate the students to their objectives, outcomes, tools, community rules and roles. The findings from this study supported incorporating the activity based game culture and collaborative culture into the classroom.

The major finding of this study which could possibly contribute to theory is that instructor support could not bring statistically significant results on the effect of digital games on learning outcomes. Keeping aside the methodological limitations like small sample size, single context study, gender sensitivity, this insignificant instructor role posits some questions and challenged to the theory. In activity theory as the instructors role has been taken as a coach, facilitator or guide which is different from the traditional teaching role. The findings of this study posits a need for theory to actually define the role of instructor in technology mediated learning environment. It raises some questions for the theory which are:

- When technology is introduced in education, does it bring a shift in teaching role?
- Do students become more independent and teachers’ role is minimized in technology mediated learning environment as pointed out by Smith (1997) that “the teacher is no longer the dominant source of information for the students?”
- Is there a difference between learning of digital immigrants and digital natives, as digital natives being more tech savvy need more autonomy for self and independent learning in technology mediated learning environment?
According to activity theory and many other previous researches, instructor support is an important factor during learning. As the game is supposed to be integrated in an instructor led classroom, hence, the level of instructor support in traditional classroom and in game session may be different. Therefore, the insignificant results in this study point out to define the difference of this role. It might be possible that the insignificant result arise due to the continuous support of instructor during game play, however, in reality, game players would ask for some freedom. The findings from this study posits a need to exactly define the level and type of instructor role in technology infused learning.

**Limitations and future research directions**

Despite the significant contribution of this study, it is not an exception from limitations. First, the study used an experimental method that was conducted in a short period. The educational settings in a university don’t allow such long and repeated measures. Hence, repeated measures and long-time investigation might provide more insights into this phenomena. Second, the study was conducted on specific subject of project management among University students. Keeping in mind that there is a severe shortage of such studies in Pakistan, this attempt is a novel approach in this country. Hence, these mentioned limitations may be acceptable. As this study focused on attitudes and opinions of students, future studies should also investigate teachers’ point of view that how much digital games have facilitated their work. Future studies are also required to study larger samples across many universities in the country which may provide deep insights into phenomena. Future studies should also account individual differences of learners and their gender on the intention to play games for education. This study is just a beginning of further practical investigation of these and many other constructs of the game based learning model in the context of developing countries, where there is scarcity of such technologies for learning and teaching. This research also demonstrated the necessity of understanding those constructs that actually bring the effectiveness of game based learning. Such as, technical attributes of games like three dimensional modelling, aesthetic effects, curriculum design and its integration in the game etc. Future researches should also investigate factors that could affect the usefulness of game play on learning efficiency such as students’ self-efficacy, population size and diversity, instructional time in class, curriculum, and teacher experience with games or technology. Hence, a longitudinal study can potentially provide profound insights into this learning process.

**Conclusion**

The experimental study conducted in context of business education demonstrated that digital game based learning approach was effective in both promoting business education students’ satisfaction and their perceived Learning Performance. It can, thus, be concluded that digital games for education are powerful tools that can be used for enhancing learning outcomes generally in higher education and specifically in business education. The study also raises some questions about the well-defined role of teacher in this experimental study of game based learning. It posits a need to determine the exact level and definition of teacher’s role in game based learning environment, as the findings of this experimental study have provided possibility of a shift in teachers’ role.

This experimental study in business education provides a hint about social change in current pedagogy setting in Pakistan. The social change is possible because the awareness and understandings of new technologies by students might increase the demand for these technologies as supplementary tools of learning and teaching. This can bring opportunity for publishers who could think about games or simulations as demanded component of textbook and curriculum material. As, it was anticipated before “…it is clear that virtual learning is an industry which is striding forward all around us…” (Blunkett, 2000).

Although not explicitly dealt with, the sample selected for this experimental study was female students. The findings of this study can be used to relate the gender based issues in technology mediated environments. In conclusion, this experimental research contributes to business education by testing the effectiveness of games as valuable supplemental pedagogy for bringing out the important learning outcomes. Hence, it is suggested to continue using as well as further exploring such innovative pedagogies for greater learning improvement in business education. This innovative pedagogy is easy to implement having strong impact on learners’ interest. To effectively educate the current business students and future managers, both research and academia settings need to do more work on pedagogical processes, inspiring students’ interest and learning effectiveness.
References


Using Mobile Learning to Support Students’ Understanding in Geometry: A Design-Based Research Study

Helen Crompton
Department of Teaching and Learning, Old Dominion University Virginia, USA // Crompton@odu.edu

(Submitted April 13, 2016 Revised October 15, 2016; Accepted November 14, 2016)

ABSTRACT
The use of mobile learning offers new affordances to teaching and learning. In this study, students from two fourth grade classes used iPads in dyads and groups to learn about angle. Using a design-based research methodology, which included observations, video, researcher journals, and artefact collection, a local instruction theory was developed on how students can learn about angle concepts through mobile learning activities. The local instruction theory is comprised of two components: (a) a seven lesson curriculum for 4th grade students on developing an early understanding of angle utilizing a mobile learning approach, and (b) additions to the scholarly theories, by providing a revised set of indicator behaviours for van Hiele levels of geometric thinking in regards to angle.

Keywords
Angles, Angle Measure, Design-based research, Geometry, Mobile learning

Introduction

Educators and governments have advocated for the use of digital technologies in classroom instruction (Bereiter & Scardamalia, 2006; Common Core State Standards Initiative, 2010). Digital technologies can be used to support mathematics teaching and learning. For example, technology offers the opportunity for students to actively participate and reorganize the way they see mathematical concepts (Stohl-Lee, Hollenbrands, & Holt-Wilson, 2010) and various mathematical representations can reveal different methods to solve problems (Heid, 2005). With technological attributes, such as the graphical capabilities, technology enhanced environments were identified for facilitating the construction of geometric understanding (Clements & Battista, 1989).

In mathematics, angles are particularly difficult concepts for students to grasp and students often develop many misconceptions and difficulties (Clements & Battista, 1989; Mitchelmore, 2002). A review of the literature reveals two strategies that appear to have been successful in supporting students with angle concepts, these are the use of Dynamic Geometry Environments (DGE; e.g., Vitale, Swart, & Black, 2014) and real-world connections (e.g., Gainsburg, 2008). DGEs provide the students with figures (e.g., lines, points, circles) and basic tools to create composite figures. Various dynamic transformations can also be performed, with the ability to trace the path of the movements for later visual inspection. Empirical evidence shows that DGEs support learning about angle as they: expand the repertoire of representations available, beyond those provided in textbooks; are a cognitive technology (Pea, 1987) acting as an external aid to amplify students’ cognitive capacities during thinking, learning, and problem solving (Lajoie & Azvedo, 2006); and provide students with a way to access the underpinning mathematical features in geometry.

There have been a number of studies to determine the affordance of teaching angle concepts with real-world connections. Researchers have used real-world objects; for example, Mitchelmore and White (2000) used adjustable models of wheels, doors, and scissors. Real-life physical situations have also been used; for instance, Fynh (2007) used a climbing project for the students to study angles made by body formations during climbing activities. Mobile learning can provide a way of bringing these two strategies (DGE and real-world connections) together. The study of students learning angles through the use of DGE has not yet been examined. Digital technologies are constantly evolving and becoming more personalized. The use of mobile technologies is becoming ubiquitous throughout today’s society. These digital technologies are also seeping into educational establishments. Mobile learning offers new affordances to teaching and learning, such as learning that is contextualized, personalized, and unrestricted by temporal and spatial constraints (Crompton, 2013), which can provide a way for students to learn about angle concepts in a more comprehensible form.

There are two research questions that guided this study:

- Are there additions to the indicators for the van Hiele levels of geometric thinking when the students are involved in mobile learning activities?
- How can mobile learning be used to facilitate students’ understanding of angle and angle measure?
Design-based research (DBR) was chosen as a method to enable the researchers to answer these two questions from the development of a local instruction theory. A local instruction theory is composed of two parts, the first part is a contribution to the theory of students learning about angle to specifically understand what additional indicator behaviours students’ exhibit when they are involved in mobile learning activities. The second part is that a mobile learning curriculum is developed for teaching angle based on the theory presented. These contributions to educational theories are important as they are highly focused towards having students learn these particular mathematical concepts. In this case, that focus is on students learning about angle.

**Theoretical framework**

**Design-based research**

Design-based research (DBR) is “a series of approaches, with the intent of producing new theories, artifacts, and practices that account for and potentially impact learning and teaching in naturalistic settings” (Barab & Squire, 2004, p. 2). The specific DBR selected for this study was developed by Gravemeijer and van Eerde (2009), as they developed a method for creating a set of exemplary instructional activities for students learning particular concepts in mathematics (Nickerson & Whitacre, 2010). Anderson and Shattuck (2012) highlighted seven characteristics of this methodology which are all used in this study. The research is; (1) situated in a real educational context, (2) focuses on the design and testing of a significant intervention, (3) uses mixed methods where appropriate, (4) involves multiple iterations, (5) involves a collaborative partnership between researchers and practitioners, (6) involves the evolution of design principles, and (7) provides practical impact on practice. Through the process of DBR, a conjectured local instruction theory is modified and strengthened.

**Geometry**

School geometry involves interlinked concepts, axiomatic representational systems and ways of reasoning that mathematize spatial objects, relationships, and transformations. Although geometry forms the foundation of learning in mathematics and other academic subjects, it is it is a difficult subject in mathematics to learn due to the abstract nature of angles (Battista, 2007; Clements & Battista, 1992) and the multiple ways in which angles can be represented (Smart, 2009). The use of digital technologies has appeared to be beneficial in mathematics; yet, mathematics teachers are often resistant to students using technologies for learning (Crompton, 2011). This opposition is often due to a lack of understanding and training in how to use technology in teaching. The mathematical theoretical framework used in this study is van Hiele’s levels of geometric thinking (van Hiele, 1984). This is used to analyse students’ geometric thought from an overarching view of shapes to a highly complex level of thinking. For example, students working at the first van Hiele level would be able to recognize a square, but could not explain why they knew that shape was a square. There are five levels altogether in the theory, but only the first three levels are used for the purpose of this study. Although the levels are not directly related to the age of the student, students of elementary age typically do not move beyond the second level of the framework. Van Hiele believed that students’ levels of geometric thought are achieved largely as a result of effective geometry instruction.

The initial indicators for each of the van Hiele levels of geometric thinking (van Hiele, 1984) were adapted by Scally (1990) and provide a criteria by which the students are matched to a particular van Hiele level of thinking in angle. The researchers in this study use these indicators to determine if they can be extended when the students are involved in mobile learning activities.

**Mobile learning**

Mobile learning is “Learning across multiple contexts, through social and content interactions, using personal electronic devices” (Crompton, 2013, p. 4). This definition includes the four central constructs of mobile learning which are learning pedagogies, technological devices, context, and social interactions (Crompton, 2013), that have been used to extend the boundaries of traditional learning. The term context refers to the subject content and the environment in which the learning takes place. Learning can take place seamlessly across multiple environments with the portability of the device. Therefore, students can learn in the real-world in which they live, connecting typically decontextualized subjects, often taught with text books, to tangible contextualized concepts.
Educators have been taking students on field trips for centuries, but in the last few years those field experiences are changing due to the technological supports that can provide additional supports to the students. Mobile devices can sense the situation of learners and provide adaptive supports to the students (Shih, Kuo, & Liu, 2012) as well as just-in-time questions, instant feedback (Hung, Hwang, Su, & Lin, 2012) and access to the Internet for further exploration of mathematical concepts. When the students are working with others, mobile computer supported collaborative learning allows for an active, motivating, dynamic environment as the device can scaffold the formation and coordination of the members of the group while the small portable form do not inhibit direct face-to-face interaction (Zurita, & Nussbaum, 2007).

Yin, Ogata, Tabata, and Yano (2010) contextualized learning as they had students using mobile device as a dynamic support in learning aspects of a foreign language. Yin et al. (2010) had the mobile device provide just-in-time scaffolding in respect of different situations in the real world. Yin and colleagues (2013) went on to develop the scaffolding participatory simulation for mobile learning (SPSML). This supportive framework adopts an experiential learning approach with five cyclical steps: the initial stage, concrete experience, observation and reflection, abstract conceptualization, and then testing in new situations. This approach is similar to the DBR approach used in this study.

Eliasson and Ramberg (2012) conducted a study that had students using a mobile software application which measured the distance between two mobile devices via Global Positioning System (GPS) to learn about area and volume. DGEs are now available on mobile devices, such as Sketchpad Explorer (2012). With this application, specific add-ons, for example, Measure a Picture (Steketee & Crompton, 2012), allow the students to interact with the real world to take photographs of physical objects in the environments and use tools within the program to measure those angles. A small number of researchers have used mobile learning to study geometry in the real world (e.g., Eliasson & Ramberg, 2012), and at this time there are none who have studied angle concepts.

Researchers have investigated students’ learning mathematics with DGE (e.g., Vitale, Swart, & Black, 2014), real-world environments (e.g., Gainsburg, 2008), and mobile devices (e.g., Shih, Kuo, & Liu, 2012). However, none have studied DGE, real-word environments, and mobile devices combined to support students learning angles. In this study, students used a DGE called Measure a Picture and that program is on an iPad. The activities took place in the school grounds. In this study, DBR is being used for its intended purpose – to bridge the gap between research and practice and provide a curriculum that teachers can use to teach angle concepts using mobile learning.

Methods

Participants

A total of 62 participants were involved in this study; two fourth grade teachers and 60 students. The fourth grade students were between 9-10 years old and 52% were males. All 60 students had participated in basic training in how to use the iPad 7 weeks before the lessons began and during that time they had also used the iPad for various applications. None of the students had used the “Measure a picture” application before this study. The participants were a convenience sample as this was one of a few schools in the Southeastern United States to have class sets of iPads. This particular grade was chosen as the Common Core State Standards requirements state that teachers should formally begin teaching angle concepts in fourth grade.

Measure a picture

As the students worked in pairs on the iPads, they primarily used an add-on program to Sketchpad Explorer. The add-on program was called Measure a Picture (Steketee & Crompton, 2012) which is a free program available on Sketch Exchange (Sketchexchange.keypress.com). The program has two measurement tools in the bottom right of the screen; the dynamic protractor and an adjustable ruler. Examples of real-world images are provided as part of the program but this image can be replaced by a photograph that is taken within the program. Real-world photographs can be taken and then the tools can be used to measure that image. Figure 1, provides a screenshot of the program with a real-world image that has two angles indicated by the placement of two dynamic protractors.
Design-based research protocol for this study

This DBR study consisted of two macro cycles with one teaching experiment occurring in each macro cycle. Each teaching experiment involved seven days of mini cycles of instructional experiments and reflection. The primary researcher acted as the teacher in both of the teaching experiments which is not uncommon in the DBR process (e.g., Cummings-Smith, 2010). The macro cycles for this study are illustrated in Figure 2. Note the occurrence of the three phases within each macro cycle: (a) the design of the instructional materials, (b) classroom-based teaching experiments and mini cycle analysis, and (c) the retrospective analysis of the teaching experiments which informed the next macro cycle.

Using the initial mobile learning curriculum developed from the review of the literature, the first teaching experiment was conducted over seven consecutive school days. The teaching experiment was the implementation of the curriculum developed in the initial stage of the macro cycle. During the teaching experiments, the co-researcher and witness observed and took notes on the classroom instruction, and the instruction was videotaped. Students’ work was collected at the end of each day. At the end of the day’s instruction, the researcher, co-researcher, and witness met to discuss the lesson and these conversations were audio recorded. Following this meeting, the researcher completed a daily reflection journal. During the retrospective analysis, in the final stage of the macro cycle, all these data were reviewed. The conjectured local instructional theory was then revised before repeating the entire process again in a different class for the second macro cycle. The local instruction theory came from the final retrospective analysis.
Data sources

One of the distinct characteristics of DBR methodology is that the researcher’s develop a deeper understanding of the phenomenon while the research is in progress. Therefore, it is essential that the research team collect a comprehensive record of the entire process (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). A list of the various data components is provided in Table 1. The horizontal headers show when these data were collected and the vertical headers are the types of data collected. These data were used in the mini cycle daily reflections and the retrospective analysis from macro cycle one and two. Scally’s (1990) angle indicators were used to code the observation notes, video, student artefacts, and researcher reflection journals.

Table 1. Data sources and when these data were analyzed

<table>
<thead>
<tr>
<th>Co-Researcher and witness classroom observations</th>
<th>Daily mini cycle analysis</th>
<th>Retrospective Analysis 1 Macro Cycle 1</th>
<th>Retrospective Analysis 2 Macro Cycle 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole-class Video</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Daily mini cycle reflection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Artifact collection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Researcher reflection journal</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Results and discussion

The purpose of this study was to answer the two research questions guiding this study. To that end, a local instruction theory was developed to extend the theory of van Hiele levels of geometric thinking (the theory) and how mobile learning can be used to facilitate students’ understanding of angle concepts (the curriculum). However, to present the results, the short mobile learning curriculum is presented first as the discussion of the theory connects with some of the activities from that curriculum.

Curriculum

A total of seven angle lessons were designed as part of this study. In Table 2 an overview of the Instructional Sequence can be found. This will give the reader some idea of the mobile learning activities and how the real-world settings were included in these lessons. The activities were a blend of mobile learning activities and in class activities to ensure that the students were able to translate what they had learned in the real-world (the contextualized) with the decontextualize activities that took place inside the classroom.

Table 2. Overview of the instructional sequence

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Learning progression</th>
<th>Instructional activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Recognize angles as geometric shapes that are formed whenever two rays share a common endpoint. Identify angles in a real-world setting.</td>
<td>Students are introduced to the concept of angle via projected images of different examples of angles in different orientations with sides of different lengths. The term angle is introduced. Students look for angles in the real-world.</td>
</tr>
<tr>
<td>2</td>
<td>Identify angles in a real-world setting. Begin to recognize that there are an infinite number of angles.</td>
<td>Students are introduced to the application Sketchpad Explorer and taught how to use the DGEs to take photographs, screenshots, and how to use the self-copying dynamic protractor. Students take photographs of angles in the playground and use the dynamic protractors to highlight the angles found.</td>
</tr>
<tr>
<td>3</td>
<td>Recognize and compare angles based on size using standard language (right, obtuse, acute, and straight angles).</td>
<td>Students introduced to the terms: right, obtuse, acute, and straight angles. Students are then involved in an activity where they use quick response (QR) codes in the playground to test to see if they are naming the angles correctly. Students then find angles in the playground and test their partner to see if they</td>
</tr>
</tbody>
</table>
4. Recognize acute, obtuse, right, and straight angles in different contexts (real-world and paper and pencil).
   Understand that angles can be measured with reference to a circle and that angles are fractions of a circle.
   Experience using a nonstandard unit of measure (a wedge).
   Recognize that the attribute being measured is the space between the two line segments caused by the turn of the line segment.

5. Recognize acute, obtuse, right, and straight angles in different contexts (real-world and paper and pencil).
   Understand that angles can be measured with reference to a circle and that angles are fractions of a circle.
   Experience using a nonstandard unit of measure (a wedge).
   Recognize that the attribute being measured is the space between the two line segments caused by the turn of the line segment.

6. Recognize that the same angle can appear to be a different size depending on different visual perspectives (positions).
   Understand that angles are defined by particular attributes which involve angle as a turn (e.g., “two rays, the common endpoint, the rotation of one ray to the other around that endpoint, and measure of that rotation”; Clements and Sarama (2009, p.186)).

Changes were made to the instructional materials throughout macro cycle 1 and macro cycle 2 from the findings during the teaching experiments and the interviews.

Summary of changes made to instructional materials in macro cycle 1

- Mathematical language reduced in Lesson One.
- Mathematical journaling was added.
- One discussion added about angles found in manufactured or natural settings.
- An infinite angle discussion was included that utilized the dynamic protractor.
- Discussion included about the importance of beginning at zero measure.

Summary of changes made to instructional materials in macro cycle 2

- After Lesson Two, the instructional plans have the teacher telling the students to just focus on one angle in their photograph, not multiple angles.
- An additional emphasis on having students discussing the mathematical concepts when working in pairs and not just point at the mobile device to explain to their partners.

Extending the theories

The additions to the van Hiele level indicators (Scally, 1990) are presented divided into five sections: drawing angles, identifying angles, sorting angles, angle measure, and angle relations. Within each section, a figure is provided in each section to articulate the revised van Hiele level indicators for angle. These revised indicators were developed from the findings of this research.

Drawing angles. Many of the students could draw figures that resembled angles, but referred to the angles by focusing on the visual characteristics. For example, one student drew a straight line angle and described the measure as about 2” as this was approximately the length of the rays. This is a common misconception that the length of the rays is salient angle attribute. Another common misconception found early in the macro cycle was that angle orientation was a salient angle attribute. For example, one student drew lines as he considered them to be angles and the lines were drawn in different directions to separate them from each other. Students worked within the visualization level of geometric thinking as they were challenged with these misconceptions.
Within the instructional sequence, students were required to consider salient and non-salient angle attributes during all of the activities. From the observations, video, and researcher reflections diary, on the last two days of instruction students did not exhibit any of these misconceptions during the activities. Furthermore, unsolicited, the students would often point out that the length of the rays and the orientation were non-salient angle attributes. As the students drew angles they would intentionally draw the rays of different lengths or draw angles in different orientations to make this point.

As the instruction began, the students often described angle categories such as right, acute and obtuse angles, but when questioned, the students did not have any further understanding beyond rote learned names and did not believe that there were different angles that could fit into those categories. Following changes and additions to the instructional plans, specific discussions were added to have students think about how many angles were in a circle and also to begin to understand that there could be a fraction of a degree. With these skills, the majority of students moved from the visualization level one to the analysis level and to indicate a possibility of drawing an infinite number of angles, this was a level of geometric thinking situated within the informal deduction level (level three) and three students demonstrated this skill in the second class.

Early in the process, another problem highlighted in through the data was that the students struggled to find words for what they wanted to describe. Students found it hard to articulate the differences between the angles they had drawn. It was a twofold issue as students did not know the vocabulary to express those meanings and also did not fully understand the mathematical concepts which made it even more difficult for the students to articulate the differences and the similarities between the angles. These skills were developed during the instructional sequence and students began to describe the angles by the degrees of measure, salient properties, and angle categories. In addition, students were also beginning to provide justifications for these descriptions. In Figure 3, the revised level indicators for drawing angles are provided. This information is based on the findings of this study as applicable to drawing angles.

**Figure 3.** Revised level indicators for drawing angles
Identifies angles. The first activity of the instructional plans had students decomposing angles into their salient components. The findings of this activity laid the foundations for understanding angle and completing the other activities in the sequence. This initial activity primarily has students working towards the analytical level two of the van Hiele levels of geometric thinking. The findings showed that a few of the students were not working at this level as they began these tasks.

Using these data collected from the two retrospective analyses, students were working within van Hiele level two. Indicators of this thinking were students’ ability to identify angles based on salient angle attributes, and to recognize non-salient attributes, such as length of the rays and orientation. In addition, students were able to generalize salient angle properties across all angles. Battista (2007) and Smart (2009) described how angles were difficult to learn due to their abstract nature and the multiple ways that angles can be represented. In the instructional sequence there were multiple opportunities provided for the students to practice using what they had learned about angle attributes to find angles in various real-world environments as well as on paper. Identifying angles in the real-world cognitive challenged the students as they had to recall and utilize this new information as they scanned across the vast amount of visual information to find the angles.

As the students searched for angles in the real world they would move back and forth between van Hiele level one and two. Initially, students used visualization (level one), as they scanned for objects that looked like angles. Students then began to determine angles by their properties moving into level two. Figure 4 provides the revised level indicators for identifying angles. This information is based on the findings of this study as applicable to identifying angles.

![Figure 4. Revised level indicators for identifying angles](image-url)
Sorts angles. The instructional plans included a number of angle sorting activities to have students consider appropriate ways to sort angles. The initial sort required students to create a set of angles and then sort those angles into groups. Students were required to consider what they had learned about angle properties and sort by those salient attributes. Many of the students did consider these properties and provided evidence of working at level two. A few of the students sorted by non-salient attributes which is indicative of thinking at van Hiele level one.

The instructional activities included angle sorts based on angle categorization. Students based those categorizations on acute, obtuse, straight and right angles. When students specifically requested further information reflex angles were also mentioned, but this additional categorizations was not utilized in the instructional activities. From the observations and video it appears that the majority of the students were able to correctly name examples of a particular category. In other words, the students were able to identify obtuse angles that were between 120° and 175°, but when the measure was 100° students would often revert to level one thinking and determine that it looked like a right angle without checking its properties.

Figure 5 provides the revised level indicators for sorting angles. This information is based on the findings of this study as applicable to sorting angles. Although students did begin to provide some justifications for sorting angles, further work is needed to have the students working within van Hiele level three of geometric thinking.

![Sorting Angles](image)

**Figure 5.** Revised level indicators for sorting angles

Angle measure. Although angle measure did not appear until later in the instructional sequence, foundational skills related to angle measure were developed from the beginning of the sequence, such as angle categorizations and discussions focused on what was being measured. In addition, the dynamic protractor was used from day two giving the students chance to build some understanding of how the turn or the ray created the angle. One of the key measurement objectives of the instructional plans was to have students internalizing benchmarks. This skill would support students in many of the measurement activities.
As expected, many of the students were working at level one at the start of instruction and this was still evident as students moved into the measurement lesson on day six. It was helpful to remind the students of certain skills they had developed as they learned about linear measurement. For example, in TE2 a discussion was added to have students remember to begin their measure at zero. This meant lining up the paper wedge along one side of the angle and looking to the other line for the measure. Another addition in TE2 was to label the 90° and 180° to support the students to internalize these benchmark measures. The revised indicators can be found in Figure 6. This information is based on the findings of this study as applicable to angle measure.

**Figure 6. Revised level indicators for angle measure**

*Angle relations.* There are two main skills encompassing angle relations, those are spatial reasoning and angle definitions. Spatial reasoning was discussed in the last section on angle measure as students struggled with spatial orientation and spatial visualization. The other skill is to understand the properties of the angles and formulate complete definitions. The instructional plans included a number of activities to have students thinking about salient and non-salient angle attributes. It also included opportunities to have students create definitions of angle.

From the observations and video, as the initial macro cycle began it was clear that some students held misconceptions to do with angle orientation. For example, one student said that a 90° angle in the orientation of an L was a left angle as it pointed towards the left. Both angle orientation and length of the rays were common misconceptions students had at the beginning of each macro cycle. Journaling was added following discussions to have the students considering what an angle is and to describe the various categories of angle. Students also had opportunities to work as a group to consider the salient properties of angle.
Over the instructional sequence, students came to understand what attributes were salient and those that were not. Students were also able to generalize properties across angles and other properties to specific angles. For example, angles between 91° to 179° are obtuse angles. Figure 7 provides the revised level indicators for angle relations. This information is based on the findings of this study as applicable to angle relations.

**Figure 7. Revised level indicators for angle relations**

In summary, the instructional sequence was effective in developing students understanding of angle and angle measure. Students in a short space of time showed significant progress across the van Hiele levels of geometric thinking in regard to drawing, identifying, and sorting angles as well as angle measure and angle relations. However, these seven days of instruction were only the initial steps and further instruction is needed if students are going to fully understand angle and angle measure. These additions to the theory on how students learn angle may be used in future research and by practitioners to develop a rich understanding of students’ understanding and use this to develop future teaching direction.

**Limitations and future research**

A limitation to this study is the initial definition of angle used in the first day of the curriculum. The angle definition developed by Clements and Sarama (2009) was designed for angles viewed in two-dimensions only. Future researchers could develop a robust three-dimensional angle definition and conduct further examination on the use of this definition with real-world mobile learning activities. Future researchers may also examine the van Hiele level of geometric thinking of pre-teachers and if programs are adequately preparing these future teachers to themselves understand angle in the way we are trying to get students to understand angle.
Conclusion and significance

In this study, the researchers used DBR to develop a local instruction theory of how 4th grade students come to understand about angle through mobile learning. Aligned to the DBR methodology, the local instruction theory is comprised of two components: (a) a mobile learning curriculum, and (b) additions to the scholarly theories. Using a cyclical iterative process of anticipation, enactment, evaluation, and revision (Gravemeijer & van Eerde, 2009) over two macro cycles, a sequence of instructional materials were developed and additions were made to van Hiele’s levels of geometric thinking (van Hiele, 1984), specifically in Scally’s (1990) angle indicators.

The activities used a mobile learning approach which had the students making real-world connections to mathematics with the use of iPads and the Sketchpad Explorer (2012) app which is a DGE. This curriculum and the apps used in this study are freely available for educators to adapt to the needs of the students in their classrooms. The curriculum can provide a springboard for educators to begin to understand the affordances of mobile learning in a mathematics classroom and develop other mobile learning activities, enabling students to contextualize mathematics to make sense of difficult concepts. The development of the level indicators are particularly useful to practitioners and researchers to go beyond students initial numerical responses to assess the students level of angle understanding in the choices they make and the vocabulary they use.

This study is significant as it appears at a time when mathematics teachers are being required to rethink their mathematical practices to go beyond the textbook to connect with real-world mathematics while using additional technological supports to extend and enhance students thinking. The promise and potential of using mobile devices is now rapidly becoming apparent and there is widespread interest amongst parents, students, principals, and teachers. One significant challenge to this implementation is the lack of teacher training and knowledge on how to successfully implement such technological tools. This study provides a list of angle indicators and a curriculum outline for learning about angles that can be used in other fourth grade classrooms.

References


Investigating the Period of Switching Roles in Pair Programming in a Primary School

Baichang Zhong1, Qiyun Wang2, Jie Chen3 and Yi Li1

1Collaborative Innovation Center for Talent Cultivating Mode in Basic Education, Nanjing Normal University, China // 2National Institute of Education, Nanyang Technological University, Singapore // 3Bixi Primary School of Changshu City, China // zhongbc@163.com // qiyun.wang@nie.edu.sg // yunizi19@163.com // yilisd@163.com

Corresponding author

(Submitted April 1, 2016; Revised November 30, 2016; Accepted February 1, 2017)

ABSTRACT

Pair programming (PP) is a useful approach to fostering computational thinking for young students. However, there are many factors impacting on the effectiveness of PP. The period of switching roles between the driver and the navigator is often ignored by researchers. Therefore, this study aimed to explore the impact of the switching period on PP. We conducted a PP experiment in four classes in the sixth grade in a primary school. The results indicated that (a) the semi-free switch was more effective for the learning achievement than the fixed periods, and the preference for adopting the fixed time interval to switch roles existed in previous studies seems to be a kind of prejudice; (b) students who switched roles every 5 minutes and semi-free were more enjoyable than those who switched roles in every task and in every class session. Moreover, the period of switching roles in every class session decreased students’ enjoyment after PP; (c) the frequency of switching roles decreased significantly, but the negotiation between the driver and the navigator became more active with time going in the semi-free class. Implications for teaching are also discussed.

Keywords

Collaborative learning, Primary education, Programmed learning, Pair programming

Introduction

Programming for K-12 can be traced back to the 1960s when Logo programming was firstly introduced as an intellectual thinking educational tool for teaching mathematics (Feurzeig, Papert, & Lawler, 2011). After Logo, the use of programming to teach thinking skills in K-12 was scarcely reported. In recent years, however, there has been renewed interest in introducing programming to K-12 students (Grover & Pea, 2013; Kafai & Burke, 2013). This was aroused by the availability of easy-to-use visual programming languages such as Scratch, Stagecast Creator and Alice, etc.

During programming, students are exposed to computational thinking (CT), a term popularized by Wing (2006). CT involves solving problems, designing systems, and understanding human behaviors, by drawing on the concepts fundamental to computer science (Wing, 2006). Many researchers argue that CT is a fundamental skill for almost everyone in a digital age, not just for computer scientists (National Research Council, 2010; Wing, 2006). More importantly, CT is in line with many 21st century competencies such as creativity, critical thinking, and problem solving (Binkley et al., 2012). Thus, it is not surprising that many educators claim that programming provides an important context and a set of opportunities for K-12 students to develop CT (Kafai & Burke, 2013; Lye & Koh, 2014; Resnick et al., 2009).

This revived interest in programming in K-12 settings suggests a need to consider how CT can be fostered effectively via programming. Studies have showed that students taught with pair programming (PP) often perform better in CT than with solo programming (Lye & Koh, 2014; Werner & Denning, 2009; Werner, Denner, Campe, & Kawamoto, 2012). PP is a practice in which two people work side-by-side at one computer, and closely collaborate to create a program. One is normally called the “driver,” who is responsible for using a computer to key in codes. The other is usually known as the “navigator” or “observer/reviewer,” who takes the responsibility for observing the driver’s work and providing support by pointing errors or offering ideas in solving a problem (Williams & Kessler, 2000).

In view of the usefulness of fostering CT, we have used PP as a pedagogical teaching technique in a primary school for two years. Meanwhile, we have also identified some issues with putting PP into practice. One main issue is about how often the roles (driver and navigator) in a pair should switch from one to the other, since it is very important to switch roles periodically between the driver and the navigator (Williams & Kessler, 2002). In other words, what period should we choose to switch the students’ roles in PP practice?
Literature review

Many studies have showed that PP has obvious benefits over solo programming, including PP can (1) significantly improve individual programming skills and promote productivity or program quality (Braithwaite, Eby, & Wahl, 2008; Cliburn, 2003; Hannay, Dybå, Arisholm, & Sjöberg, 2009; Williams & Kessler, 2000); (2) reduce frustration experienced by novice programmers; increase student satisfaction, enjoyment; and foster positive attitudes in programming (Bishop-Clark, Courte, Evans, & Howard, 2006; DeClue, 2003; LeJeune, 2006; McDowell, Werner, Bullock, & Fernald, 2002; Preston, 2005; Werner, Bullock, & Fernald, 2006); (3) increase retention of students (especially for female students) in computer science courses (Li, Plaue, & Kraemer, 2013; McDowell et al., 2006); and (4) better prepare students to work as a team (Cliburn, 2003; Williams & Kessler, 2000).

However, the above benefits do not occur automatically. Some experiments and empirical studies have reported inconclusive or contradictory results (Balijepally, Mahapatra, Nerur, & Price, 2009; Sfetsos, Stamelos, Angelis, & Deligiannis, 2009). This accentuates the need for further studies. Some factors have been identified that influence the effects of PP include:

- Task complexity (Arisholm, Gallis, Dybå, & Sjoberg, 2007; Hannay, Arisholm, Engvik, & Sjoberg, 2010);
- Partners’ skills and experiences (Hannay et al., 2010; Lui & Chan, 2006);
- Partners’ learning styles (Salleh, Mendes, & Grundy, 2011; Williams et al., 2006); and
- Partners’ personalities and temperaments (Hannay et al., 2010; Katira et al., 2004; Sfetsos et al., 2009).
- Social factors (Choi, 2015; Lewis, 2011; Zhong, Wang, & Chen, 2016)

We found that most of these empirical studies were concerned with space factors (task types and pair formations) only, but largely ignoring the time factor (period of switching roles). This one-sided consideration illustrated an incomplete picture of PP.

Many studies conducted in primary schools and higher education environments just described the period of switching roles adopted, but did not state reasons behind, and excluded it from the experiments’ variables system. For examples, Bevan, Werner, and McDowell (2002) conducted a PP experiment in a freshman programming class at UCSC, in which students alternated between driving and navigating at intervals of no more than 1 hour. Lewis (2011) conducted a study to investigate differences between PP and collaborative learning in two summer enrichment classes for students entering the sixth grade, in which the students switched their roles every 5 minutes. Mendes, Alfakhri, and Luxton-Reilly (2005) conducted a PP experiment at the University of Auckland (NZ) involved 300 second year computer science students. During each of the 90-minute lab sessions pairs swapped roles every 20 minutes, reminded by teaching assistants. Salleh, Mendes, Grundy, and Burch (2010) conducted a formal experiment at the University of Auckland to investigate the influence of personality differences among paired students. Students worked with their partners for an initial period of 30 minutes; and then swapped roles every 15-20 minutes. Salleh, Mendes, and Grundy (2014) conducted five formal experiments at the University of Auckland to investigate the effects of personality composition on PP’s effectiveness, where students worked with their partners for an initial period of 30 minutes.

Some studies took place in commercial and industrial environments with professional programmers in a flexible time interval, and investigated the frequency of switching roles. Rostaher and Hericko (2002) reported that the switching frequency correlated with programming experience and that more experienced pairs switched more often. Chong and Hurlbutt (2007) concurred that frequent switch helped the developers to maintain a high level of mutual awareness of each other’s action. Plonka, Segal, Sharp, and Linden (2011) also found that most pairs switched roles frequently and that the frequency and fluidity of switching roles indicated a high level of engagement on both developers.

These studies provide insights into the effects of switching, which indicate the switching frequency (namely, the period of switching roles) could impact on PP and should be investigated in more detail. We hypothesize that different periods of switching roles may lead to different results. Therefore, this paper reports on an experimental investigation of the time factor on PP with an expectation of generating better effectiveness in programming.

Research questions

This study aimed firstly to explore the impact of the switching period on PP effectiveness, measured by the learning achievement in PP, and learning attitude including confidence, enjoyment and value toward
programming. Our research purpose was to compare the learning achievement and attitude in different periods of switching roles. Therefore, our first two research questions were:

RQ1: Does the learning achievement vary significantly in various periods of switching roles?

RQ2: Does the learning attitude including programming confidence, enjoyment and value vary significantly in various periods of switching roles?

The second aim of this study was to compare the frequency of switching roles among three learning stages (theme 1, theme 2, and theme 3) in the semi-free switching class, and to generate implications for teaching. Our third research question was:

RQ3: Does the frequency of switching roles increase with the growth of PP time and experience?

Participants

Participants were sixth grade pupils coming from a primary school in China (it is called “school C” in this paper). We selected randomly 4 of 8 classes in the sixth grade and 150 pupils to participate in the experiment (see Table 1). These students took the Scratch introductory course in the 2014 autumn semester and Scratch advanced course in the 2015 spring semester before this study, and had basic knowledge and skills about programming. We conducted a prior knowledge quiz before this experiment, and the result of a one-way ANOVA showed that there was no significant difference among the four classes on the basic programming knowledge and skills, \( F(3, 146) = .63, p > .05. \)

We designed four various periods for Classes A, B, C, and D, respectively:
- Class A rotated roles every 5 minutes. Pairs must switch their roles every 5 minutes during PP.
- Class B rotated roles in every class session. Pairs switched their roles at the beginning of a class session which has about 40 minutes, and 20 to 30 minutes for PP.
- Class C rotated roles in every task. Pairs switched roles once in every two consecutive tasks. Each task needed about 5 to 15 minutes for PP.
- Class D rotated roles in a semi-free form. Pairs switched their roles according to their own needs, and they must switch roles at least once in a class session.

<table>
<thead>
<tr>
<th>Table 1. Sample characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Procedure

We conducted a PP teaching experiment in the 2015 autumn semester. The experiment lasted for 13 weeks in all, including 4 weeks for learning basic knowledge and skills of Alice programming, 6 weeks for formal experimental treatment. Also there were a quiz and a pretest at the beginning of the experiment and 3 achievement tests and an attitude survey at the end of the experiment (see Table 2). Every class had an additional session to learn programming each week taught by the same teacher.

<table>
<thead>
<tr>
<th>Table 2. Schedule of the experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage</td>
</tr>
<tr>
<td>Pre-experiment</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Experimental treatment</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Post-experiment</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

In the first week, the participants took part in a prior knowledge quiz to confirm that the four classes had the same level of basic knowledge and skills about programming. We also conducted an attitude survey about programming as a pretest. The participants were given a brief introduction to PP terminology and guidance prior to the Alice programming course.

In the 2<sup>nd</sup> to 4<sup>th</sup> weeks, the participants learned the basic knowledge and skills about how to programming in Alice.

In the 5<sup>th</sup> to 10<sup>th</sup> weeks, the formal PP was implemented. The participants learned to complete three programming themes in a way of PP. During the PP process, each pair exchanged the role of being a driver or being a navigator according to Table 1. Meanwhile, the teacher also used a three-stage learning progression model called Use-Modify-Create to describe a pattern of engagement (Lee et al., 2011).

In the 11<sup>th</sup> to 13<sup>th</sup> weeks, we conducted the posttests including three achievement tests and one attitude survey.

**Materials**

**Alice**

The 3D programming language Alice2.4 developed by Carnegie Mellon was used in this curriculum. Alice (http://www.alice.org/) is an easy-to-learn environment which allows users to build 3D virtual worlds.

**E-textbook**

For this experiment, we developed a school-based curriculum “learning to storytelling by programming” based on the three-dimension framework of computational thinking (Zhong, Wang, & Chen, 2016). For the convenience of improvement, the curriculum was developed into an e-textbook via 3DPageFlip (see Figure 1).

![Figure 1. A snapshot of the e-textbook](image)
Handbook for PP

For effectively switching roles in PP, we designed a student handbook, which included pair information, PP guidance, a sheet for the progress of PP (see Table 3), and a task design table. The handbook was delivered before each class session and collected back at the end of each class session. The teacher checked the handbook carefully each time, and approached the students individually if they did not fill the handbook in detail.

<table>
<thead>
<tr>
<th>Switch</th>
<th>Roles</th>
<th>Times (Minutes)</th>
<th>Task Completion</th>
<th>Who Initiated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Driver</td>
<td>Navigator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Driver</td>
<td>Navigator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Driver</td>
<td>Navigator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Driver</td>
<td>Navigator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Driver</td>
<td>Navigator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Driver</td>
<td>Navigator</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Sheet for the progress of PP

Note. This sheet must be filled in class by the navigator and calculated for a theme (a duration of two class sessions). *There were at least 10 minutes for teaching, and a maximum of 6 switches in a class session. *Filling the table cell with “driver,” “navigator,” or “consensus.” The “consensus” means the switch was not individually initiated by the drive or the navigator, but reached a consensus through negotiation between the driver and the navigator.

Measures

A single-factor experiment was employed to examine the impacts of the switching period on PP (see Figure 2). The instruments for the pretest and posttest are shown in Table 4. All students completed the pretest and posttest individually.

Figure 2. Overview of experimental design

Table 4. Instruments for data collection

<table>
<thead>
<tr>
<th>Construct</th>
<th>Source</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Knowledge (11 questions)</td>
<td>Designed by ourselves</td>
<td>.73</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Attitude Confidence (5 questions)</td>
<td>Bishop-Clark et al., 2006</td>
<td>.78</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>Enjoyment (5 questions)</td>
<td>Bishop-Clark et al., 2006</td>
<td>.75</td>
<td>.85</td>
<td></td>
</tr>
<tr>
<td>Value (4 questions)</td>
<td>Designed by ourselves</td>
<td>.76</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>Achievement (5 tasks)</td>
<td>Zhong et al., 2015</td>
<td>—</td>
<td>.86</td>
<td></td>
</tr>
</tbody>
</table>

Note. The Cronbach’s Alpha was used to check the reliability of Attitude Test and Achievement Test, the Kuder-Richardson 20 (KR-20), however, was used in Prior Knowledge Test since it was a 0-1 academic success test.
Prior knowledge quiz

The prior knowledge quiz included 11 multiple choice questions was used to measure the students’ understanding of programming concepts, since they all had taken two Scratch courses before the experiment. For instance, one question read “In order to write one piece of code to draw a square with different length of side, we can set the length of side as a: (a) variable, (b) constant, (c) parameter, (d) instruction.” The mean scores of the quiz in the four classes had no significant difference. The validity of the quiz was determined by experts’ evaluation, and the reliability coefficient of the quiz was .72.

Attitude survey

The attitude survey included three dimensions to assess participants’ confidence, enjoyment, and value toward programming. A five-point Likert scale was used, ranging from 5 = strongly agree to 1 = strongly disagree. Table 5 shows the questions used for the confidence dimension, enjoyment dimension, and value dimension.

Table 5. Attitude measures

<table>
<thead>
<tr>
<th>Type</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td>Generally I have felt secure about computer programming.</td>
</tr>
<tr>
<td>(5 questions)</td>
<td>I am sure that I could learn programming.</td>
</tr>
<tr>
<td></td>
<td>I have a lot of self-confidence when it comes to programming.</td>
</tr>
<tr>
<td></td>
<td>I am not good at programming.</td>
</tr>
<tr>
<td></td>
<td>I am not the type to do well at programming.</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>I like writing computer programs.</td>
</tr>
<tr>
<td>(5 questions)</td>
<td>Programming is enjoyable and stimulating.</td>
</tr>
<tr>
<td></td>
<td>Once I start trying to work on a program, I find it hard to stop.</td>
</tr>
<tr>
<td></td>
<td>The challenge of programming problems does not appeal to me.</td>
</tr>
<tr>
<td></td>
<td>Programming is boring.</td>
</tr>
<tr>
<td>Value</td>
<td>Programming may improve my computer skills.</td>
</tr>
<tr>
<td>(4 questions)</td>
<td>Programming is helpful for studying other subjects’ knowledge.</td>
</tr>
<tr>
<td></td>
<td>Programming makes me work in a logical and rational way.</td>
</tr>
<tr>
<td></td>
<td>Programming is helpful for solving the problems in life.</td>
</tr>
</tbody>
</table>

Achievement test

The achievement test was conducted after the PP course to assess the participants’ programming knowledge and capability. Each participant finished the achievement test individually. The test tasks consisted of five tasks adopted from a previous study by Zhong, Wang, Chen, and Li (2015) and covered two closed tasks with a defined outcome and a defined process solely, two semi-open tasks with a defined outcome and an undefined process, and one open task with an open outcome and an open process (see Table 6).

Tasks 1 to 4 had the same story context: Two hungry rabbits, one is big, the other is small. They come to a beautiful garden. They see the cauliflowers in the garden . . . (see Figure 3).

Tasks 5 had a subsequent story context: Two hungry rabbits, one is big, the other is small. They come to a beautiful garden with cauliflowers. The small rabbit eats a cauliflower and becomes weaker because she is poisoned. She cries for help . . . (see Figure 4).

Table 6. Introduction to tasks 1 to 5

<table>
<thead>
<tr>
<th>Type</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed task</td>
<td>Task1: Make the small rabbit eats off a green cauliflower.</td>
</tr>
<tr>
<td></td>
<td>Task2: The small rabbit does not stop when it eats the cauliflower, please correct it.</td>
</tr>
<tr>
<td>Semi-open task</td>
<td>Task3: Make the big rabbit jumps to the front of red cauliflowers.</td>
</tr>
<tr>
<td></td>
<td>Task4: The big rabbit moves to the front of red cauliflowers directly without jumping and swing, please correct it.</td>
</tr>
<tr>
<td>Open task</td>
<td>Task 5: Design a scenario to describe what is probably happened after the small rabbit became smaller, and need to fill out a creative design report before working.</td>
</tr>
</tbody>
</table>
The closed tasks and semi-open tasks were worth 5 points each, and the open task was worth 20 points. All together there were 40 points in total. The validity of the test was determined by experts, and the reliability coefficient of the test was .86.

All participants were also asked to provide demographic data including gender, class, and group number. The participants’ responses were coded to uniquely identify them so that the data could be further compared.

**Results**

**Achievement of programming**

The first research question was about whether there was variation in the learning achievement in PP among the different periods of switching roles. To do so, we conducted a one-way ANOVA on the learning achievement, with period as a between-groups factor.

The period of switching roles had an impact on the achievement of student programming. The one-way ANOVA demonstrated that these differences were statistically reliable, \( F(3, 146) = 2.97, p < .05, \eta^2 = .086 \) (see Table 7). The post hoc comparisons showed that Class D who switched roles in a semi-free form got higher scores (\( M = 30.22 \)) than Class A (\( M = 25.56 \)), Class C (\( M = 26.71 \)) and Class B (\( M = 27.26 \)) in the achievement test (see Table 8).
Table 7. Summary for One-Way ANOVA on achievement

<table>
<thead>
<tr>
<th>Source</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>Between-Subjects</td>
<td>422.29</td>
<td>3</td>
<td>140.76</td>
<td>2.97</td>
<td>.034*</td>
<td>.086</td>
</tr>
<tr>
<td>Within-Subjects</td>
<td>6916.54</td>
<td>146</td>
<td>47.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7338.83</td>
<td>149</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * $p < .05$.

Table 8. SNK Post Hoc Test on achievement

<table>
<thead>
<tr>
<th>Class (period of switching roles)</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A (Every 5 minutes)</td>
<td>34</td>
<td>25.56</td>
<td></td>
</tr>
<tr>
<td>Class C (Every task)</td>
<td>42</td>
<td>26.71</td>
<td></td>
</tr>
<tr>
<td>Class B (Every class session)</td>
<td>38</td>
<td>27.26</td>
<td></td>
</tr>
<tr>
<td>Class D (semi-free)</td>
<td>36</td>
<td></td>
<td>30.22</td>
</tr>
<tr>
<td>Sig.</td>
<td></td>
<td>.535</td>
<td>.074</td>
</tr>
</tbody>
</table>

Note. * $p < .05$.

Attitude of programming

The second research question was concerned with whether there was a variation in the learning attitude of programming among the different periods of switching roles. To investigate whether the learning attitude varied, we conducted a multivariate analysis of variance (MANOVA) with three dimensions of learning attitude as dependent variables and the period of switching roles as a between-subjects factor. Table 9 presents the mean scores of attitudes per period.

Table 9. Mean scores of the three attitudes in four periods of switching roles

<table>
<thead>
<tr>
<th>Class (period of switching roles)</th>
<th>Confidence</th>
<th>Enjoyment</th>
<th>Value</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class B (Every class session)</td>
<td>19.31(.54)</td>
<td>18.95(.61)</td>
<td>19.90(.55)</td>
<td>38</td>
</tr>
<tr>
<td>Class C (Every task)</td>
<td>19.11(.52)</td>
<td>19.90(.58)</td>
<td>19.91(.52)</td>
<td>42</td>
</tr>
<tr>
<td>Class D (semi-free)</td>
<td>19.91(.56)</td>
<td>20.75(.62)</td>
<td>20.19(.56)</td>
<td>36</td>
</tr>
<tr>
<td>Class A (Every 5 minutes)</td>
<td>20.29(.57)</td>
<td>21.97(.64)</td>
<td>21.41(.58)</td>
<td>34</td>
</tr>
<tr>
<td>Whole sample</td>
<td>19.63(3.34)</td>
<td>20.33(3.86)</td>
<td>20.31(3.40)</td>
<td>150</td>
</tr>
</tbody>
</table>

Note. Table entries include means and (standard deviations).

Tending to the predicted direction, the results of MANOVA, Wilks’ $\Lambda = .989$, $p < .05$, showed that periods of switching roles resulted with difference in some dimensions of attitude.

To find the difference in detail, we conducted a one-way ANOVA for each dimension of attitude. The results determined that only the mean of enjoyment was significantly different, $F(3, 146) = 4.26$, $p < .01$, $\eta^2 = .093$ (see Table 10). The post hoc comparison with SNK shows that Class A ($M = 21.97$) and Class D ($M = 20.75$) got higher scores than Class C ($M = 19.90$) and Class B ($M = 18.95$) in the dimension of enjoyment (see Table 11).

Table 10. Summary for one-way ANOVA on achievement

<table>
<thead>
<tr>
<th>Source</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>Between-Subjects</td>
<td>Confidence</td>
<td>32.67</td>
<td>3</td>
<td>10.89</td>
<td>.974</td>
<td>.020</td>
</tr>
<tr>
<td></td>
<td>Enjoyment</td>
<td>178.10</td>
<td>3</td>
<td>59.37</td>
<td>4.26</td>
<td>.006**</td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>55.20</td>
<td>3</td>
<td>18.40</td>
<td>1.61</td>
<td>.189</td>
</tr>
<tr>
<td>Within-Subjects</td>
<td>Confidence</td>
<td>1632.42</td>
<td>146</td>
<td>11.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enjoyment</td>
<td>2037.23</td>
<td>146</td>
<td>13.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>1665.07</td>
<td>146</td>
<td>11.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Confidence</td>
<td>1665.09</td>
<td>149</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enjoyment</td>
<td>2215.33</td>
<td>149</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>1720.27</td>
<td>149</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ** $p < .01$.

The reasons why the switch in a semi-free model was more enjoyable than others are rather obvious, since students’ right of learning freedom was respected. But why was the short period of 5 minutes still enjoyable? We
interviewed the students in Class A who switched roles every 5 minutes. Many students indicated that the switch was fun, as it looked like a game of carousel pattern, or seesaw.

Table 11. SNK Post Hoc Test for enjoyment (posttest)

<table>
<thead>
<tr>
<th>Class (period of switching roles)</th>
<th>N</th>
<th>Subset 1</th>
<th>Subset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class B (Every class session)</td>
<td>38</td>
<td>18.95</td>
<td></td>
</tr>
<tr>
<td>Class C (Every task)</td>
<td>42</td>
<td>19.90</td>
<td></td>
</tr>
<tr>
<td>Class D (semi-free)</td>
<td>36</td>
<td>20.75</td>
<td></td>
</tr>
<tr>
<td>Class A (Every 5 minutes)</td>
<td>34</td>
<td>21.97</td>
<td></td>
</tr>
<tr>
<td>Sig.</td>
<td>.097</td>
<td>.160</td>
<td></td>
</tr>
</tbody>
</table>

Note. "p < .05.

We conducted a supplementary paired-samples T test to reveal the difference of attitude between the pretest and posttest in the four classes. Table 12 shows, overall, the programming confidence, enjoyment and value became significantly more positive after PP than before. Interestingly, however, students who switched roles every class session actually had a decreased enjoyment level after PP ($M = 18.95$) than before ($M = 19.32$), which further indicated that the period of switching roles every class session was less enjoyable than others.

Table 12. Difference of attitude between pretest and posttest in the four classes

<table>
<thead>
<tr>
<th>Attitude</th>
<th>Class</th>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Sig.</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td>A</td>
<td>Pretest</td>
<td>34</td>
<td>17.24</td>
<td>3.25</td>
<td>-3.78</td>
<td>.001**</td>
<td>.202</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>34</td>
<td>20.29</td>
<td>2.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Pretest</td>
<td>38</td>
<td>17.61</td>
<td>3.37</td>
<td>-2.19</td>
<td>.035*</td>
<td>.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>38</td>
<td>19.32</td>
<td>3.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Pretest</td>
<td>42</td>
<td>17.02</td>
<td>3.33</td>
<td>-2.85</td>
<td>.007**</td>
<td>.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>42</td>
<td>19.12</td>
<td>3.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Pretest</td>
<td>36</td>
<td>16.44</td>
<td>3.53</td>
<td>-3.96</td>
<td>.000***</td>
<td>.214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>36</td>
<td>19.92</td>
<td>3.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td>A</td>
<td>Pretest</td>
<td>34</td>
<td>19.18</td>
<td>3.17</td>
<td>-3.66</td>
<td>.001**</td>
<td>.175</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>34</td>
<td>21.97</td>
<td>2.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Pretest</td>
<td>38</td>
<td>19.32</td>
<td>3.53</td>
<td>.389</td>
<td>.70</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>38</td>
<td>18.95</td>
<td>4.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Pretest</td>
<td>42</td>
<td>18.45</td>
<td>2.53</td>
<td>-2.42</td>
<td>.020</td>
<td>.060</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>42</td>
<td>19.90</td>
<td>3.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Pretest</td>
<td>36</td>
<td>16.92</td>
<td>4.75</td>
<td>-4.10</td>
<td>.000***</td>
<td>.168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>36</td>
<td>20.75</td>
<td>3.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td>A</td>
<td>Pretest</td>
<td>34</td>
<td>18.74</td>
<td>2.54</td>
<td>-3.90</td>
<td>.000***</td>
<td>.225</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>34</td>
<td>21.41</td>
<td>2.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Pretest</td>
<td>38</td>
<td>17.26</td>
<td>3.18</td>
<td>-3.23</td>
<td>.003**</td>
<td>.119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>38</td>
<td>19.89</td>
<td>4.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Pretest</td>
<td>42</td>
<td>16.55</td>
<td>3.05</td>
<td>-4.763</td>
<td>.000***</td>
<td>.224</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>42</td>
<td>19.90</td>
<td>3.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Pretest</td>
<td>36</td>
<td>16.50</td>
<td>3.68</td>
<td>-4.27</td>
<td>.000***</td>
<td>.216</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posttest</td>
<td>36</td>
<td>20.19</td>
<td>3.46</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. "p < .05; **p < .01; ***p < .001.

Frequency of switching roles

The third research question regarded whether the frequency of switching roles increased with the growth of PP time and experience in the semi-free class. To explore it, we calculated the Class D’s switching frequency in each theme of PP. Moreover, we calculated respectively the frequency initiated by the driver, navigator or in a consensus basis according to the records in Table 4. Certainly, we were very concerned about the switching frequency initiated by consensus and its ratio to whole, since the consensus through negotiation indicated a high level of compatibility and experience of PP.

The result (Figure 5), surprisingly, shows that the frequency of switching roles decreased significantly followed by the variations of themes, which meant more experienced pairs switched less. This decrease was confirmed by
classroom observation. Some of the dialogues inferred that the switches were initiated in a simple and crude pattern at the beginning of PP. Those dialogues read, “You do it for a long time, it’s my turn.” “I can do it, let me try!” “You cannot correct it, but I can.” “I said that we should do like this, you wait and see.” Obviously, these dialogues aimed to switch quickly by blaming. However, we heard more discussion in the third theme like “Take it easy, try not to …” “Maybe you can try it again.” “You may correct the error by changing …” which was essentially a type of dialogue called problem-solving oriented, but not switching-roles oriented.

In addition, the switching frequency initiated in a consensus basis and its ratio to whole also somewhat confirmed this decrease. There were more switches initiated by consensus, followed by the variations of themes (Theme 1: 42%, Theme 2: 50%, and Theme 3: 58%).

In conclusion, the switches of driver/navigator roles were reduced with the progress of the PP course. Nevertheless, the partners’ negotiation became more active than before.

Discussion

The period of switching roles in many PP studies generally was a fixed time interval, such as 5 minutes (Lewis, 2011), 20 minutes (Mendes et al., 2005), 15-20 minutes (Salleh et al., 2010), and 30 minutes (Salleh et al., 2014). These studies had a pre-hypothesis that the fixed time interval was the best option. There were, however, no evidence to confirm it in those studies. On the contrary, this study showed that students switching roles in a semi-free form got higher achievement than the fixed time interval from 5 to 30 minutes. The result indicated that the self-directed switch under proper pressure was more effective for the learning achievement in PP. A main reason for this result is that the right of learning freedom was respected in the semi-free switching model.

Regarding the attitude of student programming, overall, the result of this study is consistent with other studies that PP could reduce frustration experienced, increase student enjoyment, and foster positive attitudes towards programming (Bishop-Clark et al., 2006; DeClue, 2003; LeJeune, 2006; McDowell et al., 2002; McDowell et al., 2006; Preston, 2005). However, the period of switching roles every class session actually decreased students’ enjoyment of programming. On the contrary, students who switched roles every 5 minutes or semi-freely were more enjoyable than those who switched roles in every task or in every class session. Some studies conducted in commercial and industrial environments found that more experienced pairs switched more often because frequent switches indicated a high level of engagement of both programmers (Chong & Hurlbutt, 2007; Plonka, et al., 2011; Rostaher & Hericko, 2002). The result of those studies seems plausible, but the result of this study contracts that the frequency of switching roles decreased significantly with time going in the semi-free class. This difference was possibly caused by two factors as follows.

On the one hand, the participants were different. In this study, the participants were pupils who learned to program, but not adults who worked by programming. This difference led to a rigid and awkward style of switching roles indicated in the partners’ dialogues above. As a result, partners, especially the navigators, were
fussed about having an equal opportunity to operate the mouse and keyboard at the beginning of PP. Subsequently, they knew that the goal of PP was to work together to resolve programming problems and when they should switch roles with the progress of the PP course. Obviously, the skilled professional programmers would not switch their roles just for the purpose of getting equal operation time even at the beginning of PP.

On the other hand, we argue that negotiation may be more important than switch. Some previous studies concerned about who initiated the switch. Bryant, Romero, and du Boulay (2005) found that switches were mostly initiated by the driver. Contrary to the literature, Plonka et al. (2011) found that switches were frequently initiated by the navigator. However, none of those studies above focused on a special switch which was not individually initiated by the driver or the navigator, but was reached via negotiation between the driver and the navigator. This kind of switch is crucial for PP because negotiation deals with the essence of collaborative learning.

In task-oriented interactions of collaborative learning, negotiation can occur on three main levels (Allwood, Nivre, & Ahlsén, 1992; Dillenbourg, & Baker, 1996): (1) communication (meaning, signification of utterances, words, etc.); (2) task (problem-solving strategies, methods, solutions, etc.); and (3) management of the interaction on above levels 1 and 2 (coordination, feedback on perception, understanding, and attitudes.). Obviously, the switch via negotiation involves the three levels according to partners’ dialogues mentioned above.

Moreover, collaborative learning implies symmetry among partners at the same negotiation level in task-oriented interaction (Dillenbourg & Baker, 1996). This opinion is confirmed by Freudenberg, Romero, and du Boulay (2007), who found that the experienced driver and navigator tended to talk in terms of the same levels of abstraction rather than working at different levels of abstraction (Hazzan & Dubinsky, 2003). As a result, they suggested that rather than the driver and navigator roles being defined by segmenting the problem space according to the level of abstraction, they were more simply defined by the additional physical and cognitive load of typing borne by the driver. That is to say, the more important factor for PP may be not the switch of roles, but the maintenance of the same levels of abstraction which rely on the effective negotiation between the driver and the navigator. The study conducted by Vanhanen and Korpi (2007) also confirmed that the driver/navigator roles were switched only 2-3 times a day, but the partners maintained active communication and hence improved the quality of tasks.

Conclusions

This study showed that the students switching roles in a semi-free form got higher achievement than the fixed time interval from 5 to 30 minutes. The result indicated that the self-directed switch under proper pressure was more effective for the learning achievement in PP than the fixed periods. The preference of tending to adopt a fixed time interval to switch roles existed in previous studies seems like a kind of prejudice.

The results also showed that only the mean to enjoyment was significantly different among the four various periods. Students who switched roles every 5 minutes or semi-freeley were more enjoyable than those who switched roles in every task and in every class session. Moreover, the period of switching roles in every class session actually got students’ enjoyment decreased after PP.

Contrary to some existing studies, this study showed that the frequency of switching roles decreased significantly with time going in the semi-free class. This difference was possible contributed by two factors. One is that the participants were pupils who were unskilled in programming and immature in collaboration in this study, not experienced adults who worked by programming as in the previous studies. The other is that negotiation is more important than switch, which is often ignored in many PP studies. Although the switch of roles was reduced with the progress of the PP course in this research, but the negotiation between the driver and the navigator became more active than before.

The following recommendations are suggested for teachers to effectively apply PP in practice. One is choosing the semi-free model for achieving the higher learning achievement and more active attitude towards programming. Meanwhile, teachers should foster students’ awareness of self-directed switch, and guide students on how and when to switch via negotiation during the process of PP.
Another is providing scaffolding tools for PP, such as the handbook for PP which included pair information, PP guidance, a sheet for the progress of PP (just like Table 3), and a task design table. Especially, the sheet for the progress of PP can help students manage switch in the self-directed pattern.

Finally, teachers should supervise and guide switch and negotiation promptly when a meaningless switch or a lack of negotiation for a long time between the driver and the navigator happens.

Acknowledgements

The research study is sponsored by the project “Collaborative Innovation Center for Talent Cultivating Mode in Basic Education, Nanjing Normal University, China,” and the Priority Academic Program Development of Jiangsu Higher Education Institutions in China.

References


