Educational Technology & Society
An International Journal

Aims and Scope

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

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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission.

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

- Architectures for Educational Technology Systems
- Computer-Mediated Communication
- Cooperative/ Collaborative Learning and Environments
- Cultural Issues in Educational System development
- Didactic/ Pedagogical Issues and Teaching/Learning Strategies
- Distance Education/ Learning
- Distance Learning Systems
- Distributed Learning Environments
- Educational Multimedia
- Evaluation, Human-Computer Interface (HCI) Issues
- Hypermedia Systems/ Applications
- Intelligent Learning/ Tutoring Environments
- Interactive Learning Environments
- Learning by Doing
- Methodologies for Development of Educational Technology Systems
- Multimedia Systems/ Applications
- Network-Based Learning Environments
- Online Education
- Simulations for Learning
- Web Based Instruction/ Training

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; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jenni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasia 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 Merril, Brigham Young University - Hawaii, USA; Marcelo Milrad, Växjö University, Sweden; Riichiro Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoshi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshiro 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-Universität of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiski 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 Fomichov, K. E. Tsolvikovski 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-Chieh Fang, National Sun Yat-sen University, Taiwan.

Technical Manager

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

Executive Peer-Reviewers

http://www.iftes.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.
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, • Job title(s), • Organisation(s), • Full contact details of ALL authors including email address, postal address, telephone and fax numbers.

The submissions should be uploaded at http://www.ifets.info/ets_journal/upload.php. In case of difficulties, please contact ets-editors@ifets.info (Subject: Submission for Educational Technology & Society journal).
# Journal of Educational Technology & Society
## Volume 20 Number 2 2017
### Table of contents

#### Full Length Articles

- **Activities with Parents on the Computer: An Ecological Framework**  
  João C. Paiva, Carla Morais and Luciano Moreira  
  1–14

- **Experiences of Advanced High School Students in Synchronous Online Recitations**  
  Greg Mayer, Jeremy Lingle and Marion Usselman  
  15–26

- **Exploring Long-term Behavior Patterns in a Book Recommendation System for Reading**  
  Tzu-Chao Chien, Zhi-Hong Chen and Tak-Wai Chan  
  27–36

- **Learning through Blogging: Students’ Perspectives in Collaborative Blog-Enhanced Learning Communities**  
  Yu-Chun Kao, Brian R. Belland and Yu-Tung Kuo  
  37–50

- **Using Visualization to Motivate Student Participation in Collaborative Online Learning Environments**  
  Sang-Hee Jin  
  51–62

- **Test Anxiety Analysis of Chinese College Students in Computer-based Spoken English Test**  
  Yang Yanxia  
  63–73

- **Integrating Dynamic Mathematics Software into Cooperative Learning Environments in Mathematics**  
  Yılmaz Zengin and Enver Tatar  
  74–88

- **Investigating the Use of the Khan Academy and Mathematics Software with a Flipped Classroom Approach in Mathematics Teaching**  
  Yılmaz Zengin  
  89–100

- **How Does Mozart’s Music Affect Children’s Reading? The Evidence from Learning Anxiety and Reading Rates with e-Books**  
  Yen-Ning Su, Chih-Chien Kao, Chia-Cheng Hsu, Lu-Chun Pan, Shu-Chen Cheng and Yueh-Min Huang  
  101–112

  Ken Nee Chee, Noraffandy Yahaya, Nor Hasniza Ibrahim and Mohamed Noor Hasan  
  113–126

- **The Effectiveness of Using Cloud-Based Cross-Device IRS to Support Classical Chinese Learning**  
  Yi-Hsuan Wang  
  127–141

- **Creating an Online Learning Community in a Flipped Classroom to Enhance EFL Learners’ Oral Proficiency**  
  Wen-Chi Vivian Wu, Jun Scott Chien Hsieh and Jie Chi Yang  
  142–157

#### Guest Editorial

- **Learning Analytics in Technology Enhanced Language Learning**  
  Yu-Ju Lan, Nian-Shing Chen and Yao-Ting Sung  
  158–160

#### Special Issue Articles

- **Visualization Analytics for Second Language Vocabulary Learning in Virtual Worlds**  
  Indy Y. T. Hsiao, Yu-Ju Lan, Chia-Ling Kao and Ping Li  
  161–175

- **To Activate English Learning: Listen and Speak in Real Life Context with an AR Featured U-Learning System**  
  Shu-Chun Ho, Sheng-Wen Hsieh, Pei-Chen Sun and Cheng-Ming Chen  
  176–187

- **The Effects of Using the Kinect Motion-sensing Interactive System to Enhance English Learning for Elementary Students**  
  Wen Fu Pan  
  188–200

- **Using Novel Word Context Measures to Predict Human Ratings of Lexical Proficiency**  
  Cynthia M. Berger, Scott Crossley and Kristopher Kyle  
  201–212
EssayCritic: Writing to Learn with a Knowledge-Based Design Critiquing System
Anders I. Mørch, Irina Engeness, Victor C. Cheng, William K. Cheung and Kelvin C. Wong
213–223

Assessing the Language of Chat for Teamwork Dialogue
Antonette Shibani, Elizabeth Koh, Vivian Lai and Kyong Jin Shim
224–237

A Multivocal Approach in the Analysis of Online Dialogue in the Language-focused Classroom in Higher Education
Gialia Messina Dahlberg
238–250

The Effects of Cognitive Styles on the Use of Hints in Academic English: A Learning Analytics Approach
Sherry Y. Chen and Chia-Chi Yeh
251–264

A Mobile Game-based English Vocabulary Practice System Based on Portfolio Analysis
Ting-Ting Wu and Yueh-Min Huang
265–277

Article Retraction
Retraction of article - Volume 19, Issue 4, pp. 190-202
278
Activities with Parents on the Computer: An Ecological Framework

João C. Paiva¹, Carla Morais¹ and Luciano Moreira²

¹CIQUP, Unit of Science Teaching, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, Rua do Campo Alegre, Porto, Portugal // ²CIQUP, Department of Informatics Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, Porto, Portugal // jcpaiva@fc.up.pt // cmorais@fc.up.pt // lucianomoreira@fe.up.pt

*Corresponding author

(Submitted June 24, 2015; Revised October 25, 2015; Accepted June 18, 2016)

ABSTRACT

This paper proposes an ecological framework “Activities with Parents on the Computer” (APC) to bridge home and school contexts by involving parents and students in digital media based assignments. An exploratory case-study was conducted based on ten parent-child dyads that engaged in an APC at home. Attitudes were assessed through a self-evaluation questionnaire. Four parent-child dyads, that showed different patterns of attitudes, were studied in depth through semi-structured interviews. The findings revealed that parent and children have mixed attitudes towards APC. The performance in the activity varied according to the kind of parental involvement in homework and individuals’ relationship with digital media. Relevant insights helped to reframe the model in order to evaluate the relations with other living elements (e.g., friends) brought by technology into the ecological microsystems (e.g., home, school). Future research should focus on the development of more accurate instruments of evaluation, the role of teachers and other community members.

Keywords

Digital media, Homework, Computer-based tasks, Education ecology, Activities with parents on the computer

Introduction

Young people spend most of their time either at home or at school. These contexts are often strange to one another and do not communicate effectively. Given that knowledge can be distributed anywhere through Internet, it is possible to develop new socio-pedagogical, technology-based strategies to bridge school and home contexts. Technology can connect school and home contexts and help parents to follow – and get involved in – their children academic development (Lewin & Luckin, 2010). But parents often need to be instructed on how to do it (Yu, Yuen, & Park, 2012); and, despite common beliefs, students need to be taught digital skills (Ng, 2012). Digital literacy – as the very contemporary nature of literacy - is deictic (Leu, Kinzer, Coiro, Castek, & Henry, 2013), and it “refers to the multiplicity of literacies associated with the use of digital technologies” (Ng, 2012, p. 1066). As if it was not enough, many students lack scientific literacy – defined as “as the ability of people to understand and critically evaluate scientific content in order to achieve their goals” (Britt, Richter, & Rouet, 2014, p. 105). Whereas society became more digital, science and technology became more transparent and unperceived (Jenkins, Purushotma, Weigel, Clinton, & Robison, 2009). Without scientific and digital literacy one can hardly be aware of the mechanisms that elude social unbalances and, most likely, will feel helpless to act upon them.

All considered, interconnecting school and home contexts through technology is not as simple as it could seem at the beginning. Since literature on bridging school and home contexts is very associated with homework, we started by reviewing the literature on it. Then, we exposed the theoretical tenets, structure and processes involved in the framework. In the following section, methods and materials were identified and described. Finally, results were showed and discussed, conclusions were summarized and future work outlined.

Lessons from research on homework

Homework assignments can be used to create productive bonds between different settings providing students and parents with structured opportunities to collaborate (Dettmers, Trautwein, Lüdtke, Kunter, & Baumert, 2010), although they have been used by educators for different purposes, e.g, personal development, punishment, etc. (Epstein & Van Voorhis, 2001). Quality homework not only helps school to be more effective, enhancing students’ achievement, but it can also help to connect schools and homes, involving parents in their children’s academic life.
Parental involvement has been the main focus of Teachers Involve Parents in Schoolwork (TIPS), activities designed by teachers with the purpose of establishing a teacher-parent partnership through which they can help the families to be up-to-date with their children's learning activities at home while becoming involved in the process (Epstein et al., 2002). When parents get involved, children do better in school, but most families need information and guidance on how to do it in a successful way (e.g., Epstein, Van Voorhis, & Batza, 2001). Figure 1 summarizes the fundamental motives (why), behaviors (what), processes (how) and outcomes (which) that underlie parental involvement in homework (Hoover-Dempsey et al., 2001).

Parents get involved in homework because they think that they should (role construction); they perceive themselves as capable of helping their child succeed in school (self-efficacy, see Bandura, 1994); and they perceive that they are invited to participate (perceptions of invitations). What does it mean to get involved in homework? Activities range from creating physical and psychological conditions for children success, to engage in homework processes and tasks or in meta-strategies in order to adjust the task demands and the child's skills. Parents influence children through modeling, reinforcement and instruction. Modeling means that children "acquire knowledge of skills, processes, concepts and personal capabilities through observation" (Hoover-Dempsey et al., 2001, p. 203). Through reinforcement children learn by associating behaviors with desired consequences while through parental instruction they learn attitudes, skills and knowledge that are directly taught by their parents (Hoover-Dempsey et al., 2001).

Figure 1. Motives “Why, What, How, Which” that underlie parental involvement in homework (based on Hoover-Dempsey et al., 2001)

Regrettably, research has only focused on what parents do and on what students gain. Moroni, Dumont, Trautwein, Niggli and Baeriswyl (2015) concluded that the quantity of parental support in homework was negatively associated with students' achievement, while the quality of homework support was found to be a good predictor of achievement. This and other examples (Karbach, Gottschling, Spengler, Hegewald, & Spinath, 2013; Jeynes, 2012; You & Nguyen, 2011) tell us much about the motives underlying parental support and their effects on children and adolescents but tell us little about how parents are affected by engaging in homework-like tasks. Little attention has been given on how programs can promote changes that may influence not only students but parents and teachers themselves. Research on parental involvement in technology-based tasks with children...
followed a similar path. Cho and Cheon (2005) investigated the relation between family context factors and children Internet usage. Findings reveal that children were more exposed to negative Internet content than their parents thought and that engaging in shared activities reduced the exposure to negative Internet content. More recently, Nikken and Haan (2015) found out that parental mediation was enhanced by positive views of digital media, presence of elder siblings, children engagement in educational games and media skill level. On the contrary, parents feel less confident if their children use social media. Also Lee and Chae (2007) were primarily concerned with parental involvement and how it affects children. From an ecological point of view, it is fundamental to investigate bidirectional effects in dyadic interactions within microsystems (Bronfenbrenner, 1979).

Thus it seems important to develop a heuristic and ecological framework in order to help teachers to take advantage of Internet affordances to accomplish disciplinary goals (Wallace, 2004) and at the same time include parents in the dynamics of their children academic development. In the next section we will propose a framework based upon a constructivist (Piaget & Inhelder, 1969), developmental (Vygotsky, 1978) and ecological (Bronfenbrenner, 1979) approach. Piagetian and Vygotskian perspectives, although different, are far from being incompatible and exploiting their commonalities can bring new insights to research (Piaget, 1962; Nicolopoulou, 1993).

Activities with parents on the computer – An ecological framework

By APC we understand pedagogical tasks – based on socially relevant disciplinary contents– adopted or designed, assigned and evaluated by teachers, aiming to promote home and school connection, parents and students collaboration, digital and domain-specific literacy skills.

Hopefully, APC will act upon teachers, students and parents leading to changes in their usual social position, since their roles and contexts are challenged, occurring what Bronfenbrenner (1979) called “ecological transition.” An ecological transition would have happened when a student, at home, for example, explains his parents a Chemistry content as if as he was playing a teacher’s role.

![Figure 2. A techno-subsystem (reproduced with permission from Johnson & Puplampu, 2008)](image)

Johnson and Puplampu (2008) added to the original ecological model a techno sub-system. The techno sub-system is included in the micro-system and it should account for “continuously increasing complexity and availability of childhood technology” (Johnson & Puplampu, 2008, para. 11). As one can see in Figure 2, it “includes child interaction with both living (e.g., peers) and nonliving (e.g., hardware) elements of communication, information, and recreation technologies in immediate or direct environments. From an ecological perspective, the techno-subsystem mediates bidirectional interaction between the child and the microsystem” (Johnson & Puplampu, 2008, para. 11).
A closer look at Figure 2 helps us to identify and conceptualize the challenges that literacy raises to the techno-subsystem. Nowadays, being connected means that one can skip the microsystem mediation and enter in new kinds of mediation and (re)mediation (Grussin, 2004). Johnson (2010a; 2010b) gathered empirical support to validate the techno-microsystem construct but the main focus of the researches was child development and scarce attention was given to parents. This is a gap that the current study tried to tackle.

As one can observe in Figure 3, we propose a knowledge-centered approach mediated by attitudes considered as “a psychological tendency, that is expressed by evaluating a particular entity with some degree of favor or disfavor” (Eagly & Chaiken, 1998, p. 269).

![Knowledge-centered Approach Mediated by Attitudes](image.png)

Teacher is represented at the upper vertex of the triangle while students and parents are represented at the lower vertexes. It should be noted that the right side represents the relationship between teacher and students (school context); the bottom side of the triangle represents the relationship between students and parents (home context); and the left side represents the relationship between educators, i.e., teacher and parents (institutional relations).

Translating the framework into Bronfenbrenner’s (1979) terminology, one can say that students and teachers share the same microsystem (i.e., school); students and parents share another microsystem (i.e., home). Students move from one context to another often in a daily basis performing very different roles, while parents and teachers meet not so often and they are strictly focused on students, which act as a carrier of messages and meanings from school to home and vice-versa. APC acts as a mesosystem, engaging school and home microsystems in dialogue.

APC are built upon the notion of “zone of proximal development,” i.e., “the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (Vygotsky, 1978, p. 86). After Bronfenbrenner (1979, p. 9), we define development as “the person’s evolving conception of the ecological environment, and his relation to it, as well as the person’s growing capacity to discover, sustain, or alter its properties.” As such, the concept is applicable not only to students but also to parents and teachers.

Ultimately, as mentioned by Cicconi (2014), in Web 2.0, Vygotsky’s traditional proposal of the more knowledgeable other (MKO) – someone with more knowledge or a higher understanding of a particular situation/task/problem than the learner – has been transformed. The fact is that the MKO varies depending on the subject and the context. Parents are naturally expected to be more knowledgeable than their children, but one may find situations, especially among families coming from deprived social milieus, where children are the one who bring new knowledge into home. This reconceptualization provides teachers with a wide range of opportunities to promote social change.
Structure and design of an APC

Similarities can be found between APC and WebQuests. Knowing how difficult it is for young students to select and organize the enormous amount of information that is nowadays available, Dodge (1997) proposes learning activities, which guide the students while browsing the Internet. Thus, a “WebQuest is an inquiry-oriented activity in which some or all of the information that learners interact with comes from resources on the internet” (Dodge, 1997).

Exploration guides are aimed at helping learners to use specific computational simulations, through a sequence of instructions, enriched with questions and challenges in order to bridge the computer program with specific pedagogical purposes (Paiva & Costa, 2010). When a more complex computer program is expected to be used in APC, such exploration guides may be included.

APC is divided in six parts to be completed either by the student alone or by the students and their parents together, as signaled by small icons at the beginning of each part (see the bottom icons of the triangle in Figure 3). Figure 4 aims at showing a schematic view of the steps involved in APC and its relation with Bronfenbrenner’s (1979) ecological model and Johnson and Puplampu’s (2008) techno-subsystem.

The six mentioned parts are explained below:

- **Invitation:** students and parents are explicitly invited to participate, and goals and the process are explained, assuming that parents are more prompt to participate if they “perceive that their involvement is wanted and expected” (Hoover-Dempsey et al., 2001, p. 206). It acts as a mesosystem, where all participants take part. It leads to the disciplinary context.

- **Disciplinary context:** a brief description of the activity’s disciplinary context, according to science, technology, society and environment (STSE) education (e.g., Zoller, 2012), is defined by teachers (school microsystem) and it aims at increasing APC perceived legitimacy and relevance, since computers and digital technologies are often associated with less demanding activities (Kolikant, 2012). An ecological transition is expected to occur when students try to carry the disciplinary content into their home.

- **Individual area:** tasks usually computer-based (techno-subsystem) are assigned only to students. Our claim is that parental support should not replace individual work but enhance it. Autonomy is tested in ecological transitions to new settings and parents should learn how to better support their children as they grow up. After this step, parents will be called to collaborate with students.

- **Collaboration area:** parents and students are expected to work together. At least one computer-based task is included in this area. Tasks are defined in order to foster desired changes and behaviors.

- **Further research area:** this area is aimed at allowing participants to extend their research (follow-up activities). It can help teachers to identify and assess each family’s resources in order to establish challenges located in the zone of proximal development (Vygotsky, 1978).
• Self-evaluation area: parents and students are asked to assess the quality of the work they have developed through a questionnaire (see next section). Self-evaluation is an important means to trigger and organize meta-reflection processes on parents and children relationship and behavior, to promote development and to help teachers to adjust and optimize further APC proposals.

Materials and methods

In the following section we described the materials and methods used in the research.

Research question

What are the attitudes of the families towards APC and what inputs do they give us about the structure and processes that APC aimed at activating?

Materials

Bearing in mind the STSE education perspective, the APC focused on a socio-scientific chemistry topic: climate changes. In the individual area, students were asked to search the web and learn about phenomena such as greenhouse gas and acids rains, after which they should explore a computational simulation to observe how the temperature varies as a consequence of greenhouse gas concentration.

In the collaboration area, the objective was to raise the awareness about the relationship among fossil fuels, the increment of greenhouse gas effect and acidity in rain. The participants should discuss the family electricity receipt, explore a new simulation to challenge their knowledge about acid, basic and neutral solutions and finally they should explore together, through a new perspective, the simulation on greenhouse effect.

In the further research area, the family was asked to expose doubts and to identify new practices that may integrate the life of each family member, in order to help ensure the sustainability of life on Earth.

The self-evaluation questionnaire included: (i) six items (six-point Likert-type scale) to evaluate attitudes towards APC (see Table 1), that aimed primarily at promoting reflection inside each dyad; (ii) open-ended questions on the advantages and disadvantages of APC, and (iii) socio-demographic questions (e.g., occupation and age) and technology use and competence perception.

<table>
<thead>
<tr>
<th>Item #</th>
<th>Full questions</th>
<th>Short designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I liked to participate in this activity.</td>
<td>Interest</td>
</tr>
<tr>
<td>2</td>
<td>My parent/guardian engaged in this activity.</td>
<td>Hetero-engagement</td>
</tr>
<tr>
<td>3</td>
<td>It was important to cooperate with my parent/guardian in this activity.</td>
<td>Collaboration</td>
</tr>
<tr>
<td>4</td>
<td>I engaged in this activity.</td>
<td>Self-engagement</td>
</tr>
<tr>
<td>5</td>
<td>During the activity, there were unpleasant moments.</td>
<td>Unpleasant</td>
</tr>
<tr>
<td>6</td>
<td>I would be available to participate in other activities like this one in the future.</td>
<td>Future participation</td>
</tr>
</tbody>
</table>

After Johnson and Puplampu (2008) presented the theoretical proposal about the techno-microsystem, empirical studies by Johnson (2010a) assessed children’s use of the Internet by parent-report. Different methods of data gathering and different samples are necessary to support or challenge the proposal. In our study we designed a semi-structured interview to collect data from each parents-child dyad together. This approach it is an important methodological contribution to study digital ecologies.

Participants

The study involved approximately twenty participants: ten 8th grade students and ten parents (i.e., ten parent-child dyads) from which eight individuals (i.e., four dyads) were interviewed in depth (see Table 2). Qualitative studies, with relatively small samples, have given relevant contributions, for instance, to the understanding of

Table 2. Dyads characteristics

<table>
<thead>
<tr>
<th>Dyad</th>
<th>Parents</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Occupation</td>
<td>Age</td>
</tr>
<tr>
<td>A</td>
<td>Finance manager</td>
<td>46</td>
</tr>
<tr>
<td>B</td>
<td>Factory worker</td>
<td>53</td>
</tr>
<tr>
<td>C</td>
<td>Surveillance</td>
<td>44</td>
</tr>
<tr>
<td>D</td>
<td>Gardener</td>
<td>39</td>
</tr>
</tbody>
</table>

Note. *Minimum 1, maximum 5.

Data gathering procedure and analysis

Students were given an envelope containing the APC and the informed consent letter. Approximately a week later, the teachers collected the envelopes and reported to the researchers who selected, based on distinctive patterns of attitudes showed in the questionnaires, and contacted the parents to be interviewed together with their children.

Four dyads were interviewed in depth by one of the authors (a psychologist) at the school facilities in different moments. The interviews, which lasted between 26 to 40 minutes, were recorded and fully transcribed.

Quantitative data was inserted in a SPSS database and descriptive statistics were obtained. The transcriptions of the interviews were inserted in NVivo and a thematic data analysis was conducted according to the questions asked and the categories extracted from the literature on homework (see Figure 1 above), ecological development (see above the definition of “change in position”) and socio-cultural theory (e.g., “more knowledgeable other”).

Results and discussion

Overall the dyads reported to have spent between half an hour and one hour in the activity. The outcome was analyzed by one of the authors (a chemistry). Students showed different performances in the individual area. Five students answered correctly to all or almost all chemistry questions. Two students answered correctly to all or almost all questions except those questions that required accessing the computational simulation. Only two students had a clearly poor performance in the individual area.

In the collaboration area, three dyads showed good levels of collaboration and research; one answered correctly but without showing indicators of joint reflection and research. Only two dyads’ performance was clearly weak. The other five dyads only answered to part of the questions. It is important to note that besides the online simulation there was also an activity that required participants to check the electricity receipt; this activity, although in line with the STSE goals, did not require digital literacy.
In further research area, participants’ answers ranged from concrete, specific suggestions (six dyads) to absent or very elusive answers (four dyads). One should mention that it is not possible to predict the quality of the cooperation among children and parents based on the individual performance. For instance, one student that performed quite well in the individual area did not show signs of cooperation with his parents. On the other hand, we also noticed that some students who showed strong signs of cooperation with their parents did not perform as well in the individual area.

Dyad A

Dyad B

Dyad C

Dyad D

Note. *Missing answer from father C.

Figure 6. Questionnaire results by dyads

The answers to the open-ended questions were submitted to a content analysis. Parents and students considered that APC promoted learning, collaboration, and knowledge about themselves, the other and their emotional relationship. Few disadvantages have been reported when compared with advantages: lack of time; technical constraints in accessing the links; and cognitive obstacles (e.g., not knowing the answer).

The overall evaluation about APC is positive as one can see in Figure 5.

The APC was perceived as pleasant (item #1: $M = 4.7; \ SD = 1.17$) and collaboration between parents and children as important (item #3: $M = 4.7; \ SD = 1.53$). Perceived hetero-commitment (item #2: $M = 4.42; \ SD = 1.43$) and self-commitment (item #4: $M = 4.4; \ SD = 0.94$) scored high, while unpleasant moments scored very low (item #5: $M = 2.55; \ SD = 1.76$). Nonetheless, the participants were only moderately available to participate in an APC in the future ($M = 3.65; \ SD = 2.03$).

Figure 6 shows the scores on the questionnaire within each dyad. In dyads A, C and D, parents and children have similar attitudes towards APC. However, in Dyad D, the participants report that they faced unpleasant moments during the activity. In Dyad B, the differences between father and daughter are more explicit. The results of each dyad are analyzed and discussed above.
Dyad A

Dyad A was formed by a middle-age, highly-educated mother, who worked as finance manager at university, and a daughter with good academic achievement. The relation between mother and daughter was positive and digital technologies were not perceived as problematic. The mother was strongly dependent on technologies in her professional daily-life but she did not use social networks. On the other hand, the daughter used Facebook with a high perception of control over her own behavior (perception that was shared with her mother).

The mother responded to her daughter homework, trying to give the structures for her daughter to develop herself, advising and instructing her. She felt more confident when homework was about mathematics (her academic expertise area), but she acknowledged that sometimes her explanations about the subject could be too elaborate for her daughter to understand.

The dyad answered to all the questions of the APC acknowledging that the activity made them reflect upon new themes (STSE) and lead them to collaborate more than they usually did in homework.

“It is good to step out of our subject area and get in touch with other subjects especially with this one which is so important related with the environment.” [Mother A]

Mother and daughter showed a positive and identical evaluation on APC (Figure 6). From a qualitative perspective, APC increased the perceived relevance of parental involvement in homework:

“If I got more involved in [her] activities she could have had more fives [equivalent to the highest grade]. This year I think that I ought to have stayed a little more with her to understand where she had more difficulties.” [Mother A]

In line with You and Nguyen (2011, p. 556) it is evident that “having a close bond and showing personal interest and care for the child does, in fact, have far-reaching consequences in the lives of children.”

Dyad B

Dyad B was formed by a 53 years-old father, who worked at a factory, and a daughter with poor academic achievement. He did not use the computer and had a low perception of competence on digital technologies. Daughter and father perceived her relation with Internet as problematic:

“She isolates herself. She spends hours and hours alone with the computer… doesn’t even eat and sometimes we don’t even notice her, she is there, alone.” [Father B]

“I see that I isolate myself, but I could be in another part of the house with the rest of the people, but being on the computer I would not be truly there…” [Daughter B]

According to literature, total time using Internet affects perceived family time but not family communication (Lee & Chae, 2007). In this case time and communication were affected. Typically, father B did not get involved in his daughter homework tasks, justifying that she arrived late at home and that she was given explanations on a daily basis elsewhere. In fact, it was the first time that they worked and studied together and both acknowledged that it was not easy to collaborate.

“She does not accept that much my opinion... she has difficulties in accepting...” [Father B]

“I think we both have different opinions, but then it was only a matter of putting all in one.” [Daughter B]

Based on the quantitative results, one could say that father and daughter developed different attitudes towards APC (Figure 6). The daughter evaluated APC more favorably than her father, although less favorably than her schoolmates. She did not report any unpleasant moment during the activity. Although the father has scored three in all questions, his attitudes and the meaning of the experience became more visible during the interview. STSE subject was highly valued by the father who examined the electricity receipt with his daughter easily and with confidence. The digital activity was led by the daughter, who introduced her father to Internet and computers, despite the fact that she was not able to access one of the computational simulation available online.
I think that it is positive. It brings people together. We had an opportunity to make more use of the computer. I, for instance, have a computer at home, but I do not use it. [Father B]

**Dyad C**

Dyad C was formed by a middle-aged father, who worked as vigilant and perceived himself as relatively competent using the computer, and a daughter with medium academic achievement who also reported being competent on digital technologies. Now and then, the mother also participated in the activity. The father usually engaged in his daughter homework, instructing her through an almost strictly cognitive approach:

“I study with her. She studies first and then she comes to me to check if it is correct. I also ask her questions, how much she has studied, if she has learnt anything or not…” [Father C]

Both of them acknowledged being easy to find information on Internet but also that it was necessary to understand it and to reproduce it with new words (avoiding copy and paste processes).

Father and daughter evaluated highly their experience with the APC (Figure 6). The daughter considered that the APC was important because parents could monitor what children learn at school. As Jeynes (2012) suggested, the efficacy of parental involvement programs depend on the quality of the parent-children interaction, monitoring included.

Father and daughter considered that APC helped to show the relevance of understanding different points of view and was efficient in bridging people’s opinions. They acknowledged that APC involved some stress (which can be taken as a sign of having experimented cognitive dissonance).

**Dyad D**

Dyad D was formed by a middle-aged mother (a gardener), and a son, with medium-low academic achievement who has experienced a problematic relationship with digital technologies in the recent past. His mother had to limit the use of the computer because he was spending too much time playing games. His father was working abroad and engaged in the APC through Skype, an unexpected answer to a contingency (geographic dispersion) that nowadays affects innumerable families.

Considering Figure 6, the evaluation is ambivalent. Although the unpleasant moments’ item scored high and the willingness to participate again scored low, one should note that the interview did not support these values.

Mother D overtly acknowledged that she was no longer capable of helping her son on a disciplinary content basis. To overcome the handicap, they had developed a strategy that was very tuned with the expected collaboration to perform APC.

“Honestly, I cannot follow the subjects anymore. Sometimes, I will check if the homework is done, but he got used to do it on his own early on.” [Mother D]

“(…) from early on I developed a tactic to make him studied: it was to pretend that I did not understand the subject – this was not the case, because I did not understand – which made him read and study, and let us say that he was delighted, because he thought he was explaining that subject. [Mother D]

I felt like if I was a teacher explaining. [Son D]

They did not answer to all the questions of APC, but their cooperation was visible and positive. Nonetheless, the intensity of the relationship was somewhat excessive, and the son seemed still dependent on his mother to control his relation with digital games (e.g., the computer was in mother’s room, so that he could not use it during the night).

**Wrapping up and reframing APC**

The digital media may become a source of tension within the family, requiring dialogic skills from parents and students (dyads B and D). Mother D limited the use of computer in time and space: if she helped her son coping
with the immediate problem or dependence, she did not provide her son with means to emancipate his relationship with digital media or to build his autonomy: “simply prohibiting or restricting seems ineffective for guiding children’s Internet use” (Lee & Chae, 2007, p. 644). Furthermore, parents’ confidence depends on how active on social media children are (Nikken & Haan, 2015) and “when parents’ control or guidance over the computer use turned out to be ineffective and their children refused to communicate with them about these issues, they became ‘worried outsiders’” (Yu, Yuen, & Park, 2012, p. 19). It is not irrelevant that family B and D had a low socio-economic status, a factor highlighted by the literature on digital divide (e.g., Ritzhaupt, Liu, Dawson, & Barron, 2013). Both home microsystems include technology but dialogue, especially in Dyad B, about technology is rather poor if not incipient.

The inability to access the web resources confirmed the need to promote digital literacy, among youth and adults (Ng, 2012). When neither parents nor students possess adequate levels of digital literacy to address specific challenges, it is up to the teacher to take advantage of how students change position within the home microsystem (Bronfenbrenner, 1979).

Personal or digital mediated communication from teachers to parents can be a good strategy to increase the visibility of the invitations. While almost all students successfully completed individual area, collaboration area was not and adjustments are necessary. Further research area helped families to bridge theory and practice, turning knowledge in behaviors that can organize family daily-life on sustainability, but did not gave us enough indicators about the families’ resources. As such, it is necessary to rewrite this section in order to map the ZPD, according to Vygotsky (1978).

![Figure 7. Remodeled APC framework](image-url)

It is not clear if changes in the process of role construction only became meaningful or not during the interview. The fact is that (i) parents’ insights (A and C) on their usual approaches to homework opened the possibility of adopting behaviors that do not require academic expertise (Simpkins, Price, & Garcia, 2015) and (ii) students assumed the MKO role in relation with digital technologies (daughter B) and in relation with disciplinary content (son D).

Children establish emotional and “academic” bonds with their parents and teachers. At the school microsystem they primarily play the role of students, interacting with their teachers while developing attitudes towards them and school. At the home microsystem, they are primarily children, interacting with their parents and relatives: school is brought into their homes through homework assignments or grades. Within a specific setting each person is also connected with other living objects of evaluation (e.g., friends) that can be activated through digital media. APC overlap schoolwork and digital media while overtly involve parents, thus creating new triadic relations inside the triangle formed by students, parents and teachers. The smaller triangles can intensify, change or disrupt previous attitudes as predicted by Heider’s (1946) balance theory, through psychological processes, e.g., cognitive dissonance (Festinger, 1957). In dyads B and C, conflict was not merely cognitive, but was also emotional, since it equalized the value of parents and students’ opinions. These insights lead us to reframe our model, in order to account for micro-relationships (see Figure 7, d, e, f). The APC framework should be Internet-
based and compatible with multiple devices, from computers to smartphones, in order to help teachers to implement participatory and geo-localized activities that increase the visibility of its socio-scientific meaning.

**Conclusion**

In this paper we have presented APC as a means to interconnect school and home and to promote literacy. The scope of this research has been mainly circumscribed to the participants’ attitudes towards APC inside each parent-child dyad.

Although the result’s generalizability is limited, small-scale studies provide powerful theoretical and practical insights. Bearing in mind an ecological framework, we observed that APC challenges parents and children to play alternative roles. Since digital media usage affect family time as well as communication, to restructure the way children and parents approach technology can increase the quality of family relationships.

Literature is mainly focused on the effects of parental involvement in homework or digital-based tasks in children academic achievement or development. Our research uncovered the possibility of parents - and not only children - be affected by shared technology-mediated interactions. Whether or not these changes only affect parental involvement models or are also manifested in other dimensions of parents’ life, there is no doubt that this topic deserves to be investigated.

Johnson (2010b) hypothesized that mesosystem lost relevance in contemporary society. On the contrary, the current findings reinforced mesosystem as a theoretical tenet with practical significance. The alternative hypothesis is that digital media are more than a nonliving object in the techno-microsystem. They are a means through which different systems communicate, one that encompasses mesosystem in a highly sophisticated society. As highlighted by one of the reviewers, school teachers, educators and principals can apply APC to foster school-parents relationships.

Future work should try to gather more qualitative data from other settings in order to suppress time and sampling limitations that affected this research. It is our choice to conduct ecologically-grounded research instead of quantitative approaches, including direct observation in order to track how many time parents and children spend on each area and assess their interactions. Nonetheless, it would be important to use more reliable quantitative assessment of some of the tenets of the framework, including data gather instruments, e.g., questionnaires on homework and on the relationship with teachers. A second line of research consists in investigating teacher’s role during the process of creating, implementing and evaluating an APC. What variables are at stake when APC is discussed inside classroom with students and in the regular meetings with parents? Future work should also discuss if APC is to be strictly aimed at parents or if it can also be aimed at other relatives or other members of the community.

**Acknowledgements**

The authors are grateful to Ana Parada for her diagram designs for the APC model, to Filomena Gaspar and Ana Mouta for their useful comments during the initial draft process, and to João Moreira who kindly proofread the paper.

Luciano Moreira is supported by the grant PD/BD/114152/2015 from the Fundação para a Ciência e a Tecnologia, IP.

**References**


Experiences of Advanced High School Students in Synchronous Online Recitations

Greg Mayer¹, Jeremy Lingle² and Marion Usselman²

¹School of Mathematics, Georgia Institute of Technology, Georgia, USA // ²The Centre for Education Integrating Science, Mathematics, & Computing, Georgia Institute of Technology, Georgia, USA // greg.mayer@gatech.edu // jeremy.lingle@ceismc.gatech.edu // marion.usselman@ceismc.gatech.edu

Corresponding author

 radicals [{Submitted September 22, 2015; Revised January 13, 2016; Accepted May 9, 2016}]

ABSTRACT
The question of how to best design an online course that promotes student-centred learning is an area of ongoing research. This mixed-methods study focused on a section of advanced high school students, in college-level mathematics courses, that used a synchronous online environment mediated over web-conferencing software, and whether the affordance of multiple communication channels and student-centred activities affected involvement, cohesion, and satisfaction. Study participants reported that anonymous input and group work activities encouraged their involvement in learning activities, increased their satisfaction, and fostered social cohesion. Although the on-going management of technical issues limited student involvement and satisfaction, there were no differences in final grades obtained by students participating in this delivery format and their peers participating in an alternate learning environment facilitated by video-teleconferencing. This study offers supporting evidence that a student-centred learning environment mediated over web conferencing software can foster social cohesion, student involvement, and student satisfaction.

Keywords
Web conferencing, Synchronous learning, High school

Introduction

Distance learning can take many formats, ranging from live but non-interactive radio broadcasts, to highly facilitated and interactive web-based courses, to completely asynchronous large scale Massive Open Online Courses. Each affords different benefits and is appropriate for different audiences, and the challenge for course designers and instructors is to best utilize the different platforms to maximize participant learning.

This study is part of the evaluation of the Georgia Tech Distance Calculus Program (DCP), which is an Advanced Calculus and Linear Algebra course sequence offered to a live audience of traditional on-campus college students and simultaneously to approximately 450 high school students via synchronous video teleconferencing (VTC). In an effort to increase a sense of community among those high school students that have exceptionally strong math skills but, who attend schools without a solid peer group of like-minded students, Georgia Tech initiated an experimental DCP recitation section that used computer-mediated communication (CMC) in a synchronous online learning environment, mediated using web conferencing (WC) software. The goal was to promote student-centered learning and encourage active student involvement and cohesion.

It is widely accepted in the distance education literature that WC software can be used to immediately provide feedback in real time and build a sense of social cohesion among students (Hrastinski, 2008; Oztok, Zingaro, Brett, & Hewitt, 2013). Unlike older forms of synchronous communication in distance education, including VTC, WC technologies can support simultaneous communication over several different channels, or media. These media include instant messaging (IM), polls, audio and video, as well as a shared whiteboard that participants may contribute to anonymously and that enables students to import, collaboratively share, and annotate various types of documents. Some WC tools allow breakout rooms where students can engage in synchronous group work. The benefits and challenges of facilitating group work in synchronous online environments has not yet been extensively studied, but in asynchronous environments recent case studies have found that small group work activities can develop teamwork skills, trust, and cognitive processes among learners (Biasutti, 2011; Tseng & Yeh, 2013).

Students and instructors communicating over multiple channels creates unique challenges in facilitating learning in ways that promote student-centered learning and social cohesion without unnecessarily introducing cognitive overload, technical issues, and off-topic conversation (Cornelius, 2014; Cornelius & Gordon, 2013; Kear, Chetwynd, Williams, & Donelan, 2012; Martin, Parker, & Deale, 2012; Olson & McCracken, 2015). The process
of managing multiple channels has been described as overwhelming and stressful to facilitators of WC learning environments (Cornelius, 2014; Kear et al., 2012; Peacock et al., 2012). To address these challenges, it has been recommended that communication be limited to only those media that are needed (Cornelius & Gordon, 2013; Martin et al., 2012). Martin et al. (2012) list teaching strategies for instructors who are new to the WC environment. They suggest that the instructor “do not give eboard access unless they need it” and that the “private chat option can be disabled if you do not see the need for it … students prefer to use the private chat option to talk to their classmates/teammates” (p. 249).

It is not yet clear which communication channels are best suited for a given learning activity in the WC environment. Some argue that audio and video channels are particularly helpful for fostering social presence (Kear et al., 2012, p. 962; Peacock et al., 2012) and social bonding (Cornelius, 2014, p. 268). On the other hand, one study found that high school students prefer to use IM over other communication channels in the WC environment (Murphy, Rodríguez-Manzanares, & Barbour, 2011), and another argued that the WC facilitator “consider whether video or audio is really necessary”, and to “choose how to use the media at your disposal to suit the situation” (Cornelius & Gordon, 2013, p. 280). Certain channels can also present more technical challenges than others. IM has been reported to be useful for those participants who are experiencing technical issues with other channels and provides a medium that supports socialization and elaboration (Carrington, Kim, & Strooper, 2010; Cornelius, 2014, p. 267). There is much to be learned in terms of identifying the learning activities, teaching practices, and communication channels that best guide learners towards a given set of learning goals.

The present study explores the following questions regarding the experiences of the high school students in the DCP.

- Is there a difference between the grades obtained by students participating in a WC delivery format, relative to their peers who participate in a delivery format administered over VTC?
- Do different learning activities and the affordance of different communication media affect student involvement, student satisfaction, and social cohesion during synchronous online learning sessions mediated with WC software?
- How do high school students use technologies to communicate with other participants in a synchronous learning environment mediated over WC software?

The first research question aims to explore whether the final grades obtained by students in the WC format are different from those of their peers who are enrolled in the same program and instead rely on VTC technology. The last two questions aim to characterize the learning activities and communication channels that students benefit from and how they may be better supported in future iterations of this distance education program.

Methods

Participants and context

The sample for this study consisted of twenty advanced high school students who, in the 2013/14 academic year, were enrolled in two consecutive multi-section mathematics courses offered through the Georgia Tech DCP (Morley, Usselman, Clark, & Baker, 2009). The first course focused primarily on Linear Algebra, the second on Multi-Variable Calculus. These courses are simultaneously offered to undergraduate students attending Georgia Tech and to high school students who were distributed throughout Georgia.

DCP courses offer synchronous 50-minute sessions five mornings per week. Students view live lectures that are facilitated by an instructor for three of these mornings, and during the other mornings students are divided into sections of roughly 50 to 60 high school students and connect to recitation sessions. In these sessions, teaching assistants (TAs) solve problems on concepts that students have encountered in lectures and assignments. TAs who facilitate these sessions are either graduate or undergraduate students and have completed a mandatory training course that deals with university teaching policies, as well as teaching strategies that promote active learning. TAs are also offered a training session on how to communicate using VTC technologies, and university staff are physically present during all of their recitations to provide technical support. All TAs are responsible for identifying and facilitating learning activities for their sessions that are aligned with course objectives and assessments.
High school students have connected to lectures and recitations through VTC since the DCP began in 2005. An additional format that relies on WC software was offered in the 2012/13 academic year for a small group of high school distance students (Mayer & Hendricks, 2014). Its primary purpose was to offer a delivery model that might promote engagement and increase community among those who were the only student in their school participating in the DCP, and possibly increase enrollment from schools that could not otherwise participate in this program due to financial barriers introduced by the use of VTC equipment.

For the first six months of the 2013/14 academic year, the WC recitation section was facilitated using Wimba Classroom, and for the last three months (for reasons that will be described later in this report) recitations were facilitated in Adobe Connect. Study participants were lent microphones and Wacom Bamboo splash tablets for the duration of the program to help them participate in recitation activities. Students communicated through a number of methods during recitations, including instant messaging (IM) that was viewable to the group as a whole, private messaging, microphones, and the whiteboard editable either through the tablets or using a computer mouse. The WC tools also accommodated group work activities through the use of breakout rooms. Students in group work activities were given a set of problems from past quizzes that they would work on in groups of three to five students for ten to fifteen minutes, after which they were discussed as a group.

All twenty participants in this study were enrolled in the WC section. When applying to the DCP, all students were asked to indicate their academic year, gender and ethnicity, but not asked for their birthdate or age. Among the twenty participants, 19 identified as high school seniors, 1 as a high school junior; 16 identified as male, 4 as female; Ethnically/racially, 11 students identified as White, 5 as Asian, 3 as having two or more racial/ethnic identities, and 1 as Latino/a.

High school students admitted to the DCP are required to meet a high level of performance in mathematics. Minimum enrollment requirements include a math GPA of at least 3.5, completion of the Advanced Placement Calculus AB or BC course earning a 4 or higher on the AB Calculus exam or a 3 or higher on the BC exam, and a score of at least 600 on the Math SAT. Meeting these minimum requirements does not guarantee acceptance due to the competitiveness of the program (Eligibility Guidelines, n.d.).

**Data collection methods**

A mixed methods approach was utilized to answer the posed research questions. Using both quantitative and qualitative data allows for triangulation of the findings where possible, and for the weaknesses of certain methods to be balanced by the strengths of others, which improves the trustworthiness of findings (Rossi, Lipsey, & Freeman, 2003).

Research artifacts included student surveys, grades, logs of online activity, and online focus group discussions. In order to determine if differences existed in the academic performance between the students participating in the new WC section and their peers in the traditional VTC sections, average grades earned by students during the 2013/14 academic year were examined. In addition, average grades of undergraduate students participating in the campus-based face-to-face sections were also recorded, though the grades were expected to be lower due to the competitive admission requirements for the DCP.

To characterize the learning environment in the WC section, a modified version of the College and University Classroom Environment Inventory (CUCEI) (Fraser & Treagust, 1986) was administered to participating students. The CUCEI is a validated survey instrument that was developed to measure seven dimensions of actual and preferred classroom environments among students taking undergraduate and post-graduate courses. Three of the constructs that captured student perception of the actual environment were selected for the current study: (1) Involvement (the extent to which students participate actively and attentively in class discussions and activities), (2) Student Cohesiveness (the extent to which students know, help, and are friendly towards each other), and (3) Satisfaction (the extent of enjoyment of classes). Each construct has seven items, and Fraser and Treagust (1986) found the Cronbach’s alpha coefficients for these constructs in their study to be 0.70, 0.90, and 0.88, respectively. The original items were modified in the present study to reflect the current research context by replacing references to “class” with “recitation” and by replacing “instructor” with “teaching assistant.” The original four-point Likert-type response scale was maintained, which ranged from “strongly disagree” to “strongly agree.” In addition, two open-ended question, “how could recitations be improved” and “if there’s anything else you’d like us to know about the course, please write it here,” were included following the rating items to provide a better understanding of students’ perspectives of the recitation section. This survey was administered in August 2013 and in April 2014.
A “Technology Survey” was administered in September 2013 and April 2014, which focused on how students used technology during their recitations. The appendix contains the entire survey. Both the modified CUCEI and the Technology Survey were administered online.

The WC tools that were used in this study allowed students to interact with each other and with the TA via IM, a shared whiteboard, and microphones. Figure 1 shows a screen capture of a moment during a whole group discussion, facilitated by Wimba Classroom. Students could also see a video feed of their TA captured with a web camera, which is redacted in the figure for the purposes of ensuring a blind review. Due to technical limitations of Wimba Classroom, students were not allowed to use web cameras.

Wimba Classroom and Adobe Connect allow students to use IM to send messages that all participants can read. Data for the current study included IM transcripts of 45 of the 49 recitations held during the 2013/14 academic year, as well as transcripts for all seven group-work activities that were conducted in Adobe Connect. Private messages sent between students cannot be recorded in either of these WC tools. Both Wimba Classroom and Adobe Connect archive the times that participants log in and out: this log report was used, along with the recitation transcripts, to calculate the rate at which students wrote comments during recitations (Martin et al., 2012; Lobel, Swedburg, & Neubauer, 2002; Oztok et al., 2013). Table 1 shows an example of the transcript data - the first 14 comments from the beginning of an exchange between four students in a group work activity.

<table>
<thead>
<tr>
<th>Comment number</th>
<th>Student code</th>
<th>Time</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>8:35</td>
<td>Hello friends</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>8:35</td>
<td>Hello</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>8:35</td>
<td>hello there</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>8:35</td>
<td>Hi there!</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>8:35</td>
<td>So x is from 0 to 2, everything else is in terms of y and z</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>8:35</td>
<td>wow that was an awful x-axis</td>
</tr>
<tr>
<td>7</td>
<td>19</td>
<td>8:37</td>
<td>What line is that? z = 2 - y?</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>8:38</td>
<td>What’s the green line?</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>8:38</td>
<td>y+z = 2</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>8:38</td>
<td>z = 2 - y</td>
</tr>
<tr>
<td>11</td>
<td>19</td>
<td>8:38</td>
<td>Okay, that’s what I thought. Thanks guys.</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>8:39</td>
<td>So let’s do dzdydx</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>8:39</td>
<td>x first cause it’s the easiest</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>8:41</td>
<td>so we need 2 triple integrals?</td>
</tr>
</tbody>
</table>

The comments in Table 1 are provided to give the reader a sense of some of the interactions that are possible with group work in this synchronous online environment. A screen capture of the whiteboard that shows the end product of their concurrent group work is shown in Figure 2.

The final source of data for this study was a set of focus group discussions held during recitations for the purpose of gaining further insight into how students preferred to interact with their peers. These discussions were held at
the end of the program (April 2014). During these sessions, students were only able to participate through the use of the IM tool that the WC software provided. Like survey participation, focus group participation was completely voluntary. The focus groups were semi-structured and included the following prompts.

- Students can write on the board at any time. In what ways, if any, did this help your learning in recitations?
- Is it important to get to know other students in recitation? Why/why not?
- Most students didn’t communicate with microphones very often. Why do you think this was the case?

In order to minimize the impact on learning activities, brief focus groups were held with the same students in three consecutive recitation sections. In each focus group, one of the three prompts was presented along with any needed follow-up questions for clarification. 12 students were present for the first question, 11 students for the second, and 13 students for the third. Identified themes from the focus groups data were triangulated with survey results in order to verify and strengthen study conclusions. All qualitative data were analyzed for thematic content in which codes were first developed using open and axial coding, which were then organized and linked to form key categories (Miles & Huberman, 1994; Neuman, 2011, p. 512).

Results

Academic achievement data

Average grades for high school students participating in the WC recitation section, in the Distance Calculus VTC sections, and undergraduates in the on-campus face-to-face sections are presented in Table 2.

<table>
<thead>
<tr>
<th>Delivery format</th>
<th>Fall 2013</th>
<th>Spring 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Average (SD)</td>
</tr>
<tr>
<td>Distance learning, WC</td>
<td>20</td>
<td>95% (0.05)</td>
</tr>
<tr>
<td>Distance learning, VTC</td>
<td>307</td>
<td>95% (0.07)</td>
</tr>
<tr>
<td>Face-to-face</td>
<td>212</td>
<td>88% (0.11)</td>
</tr>
</tbody>
</table>

In both semesters, differences in final grades obtained by students in the WC section were not statistically significant from those in the VTC section, suggesting that students participating in these two formats performed at academically similar levels. Final grades obtained by the on-campus students (the face-to-face delivery model) are included for comparative purposes and show lower average grades than those among students in the DCP. This finding is expected given the stringent admission criteria in mathematics for those in this program, and the fact that participating students all needed to have successfully completed Advanced Placement Calculus before their senior year in high school (Eligibility Guidelines, n.d.). DCP students are drawn from a pool that contains only the most academically advanced students in the state.
Student involvement, satisfaction, and cohesion

The modified items from the CUCEI were administered to the 20 students participating in the WC recitation section. Responses were converted to numerical values (strongly disagree = 1, disagree = 2, agree = 3, strongly agree = 4). Measures of internal consistency for each of the three constructs, Cohesiveness, Involvement, and Satisfaction, were calculated using Cronbach’s alpha and were found to be adequate (Kline, 2005) at 0.69, 0.77, and 0.71, respectively.

Items within each construct were averaged and then a grand average calculated across all participants. As noted previously, the instrument was administered at the beginning and end of the 2013/14 school year. Of the 20 members of the recitation section, 10 students (50%) provided responses at both time points. Table 3 presents the average ratings per construct for these students.

<table>
<thead>
<tr>
<th>Construct</th>
<th>August 2013 (SD)</th>
<th>April 2014 (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohesiveness</td>
<td>2.08 (0.32)</td>
<td>2.52 (0.40)</td>
</tr>
<tr>
<td>Involvement</td>
<td>3.20 (0.28)</td>
<td>3.28 (0.37)</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>3.26 (0.41)</td>
<td>3.44 (0.43)</td>
</tr>
</tbody>
</table>

Satisfaction and Involvement scores were above the midpoint of 2.5 (3.26 and 3.20, respectively) at the August 2013 administration. The Cohesiveness construct was below the midpoint, with an average of 2.08 overall (near “disagree” = 2). At the conclusion of the two semesters, each of the three constructs showed an increase. Considering the matched students, the largest changes occurred in the measure of Cohesiveness (increase of 0.44), which was the only construct among the three found to have a significant increase between the August 2013 and April 2014 data ($t(9) = 4.99$, $p < .01$).

Students had high levels of Satisfaction and Involvement early in the school year and these constructs increased non-significantly by the end of the program. Students began the school year with low levels of Cohesiveness, as would be expected because they did not previously know one another, but their cohesion increased by the conclusion of the year. Because non-significant results may be due to lack of statistical power due to the small sample size, a post hoc power analysis was conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) with power set to 0.80 and alpha = 0.05. Given the effect size for Satisfaction of 0.43 and for Involvement of 0.24, the analyses indicated that sample sizes would need to increase to 47 and 136, respectively, in order to detect group mean differences.

In order to better understand these results, responses to open-ended questions in the surveys, and focus group data were examined. With respect to social cohesion, a focus group session held at the end of the program explored the benefits of getting to know other students. Three students (among eleven present) described how getting to know other students had a positive impact on their learning through increasing their level of satisfaction with their recitations. One of those three students wrote “It makes me look forward to recitations even more.” Four students in the August CUCEI survey indicated that group work was an activity that helped them develop cohesion. But other students wrote that they did not feel that cohesion was necessary. One student wrote in the April CUCEI survey, “I don’t really think it’s necessary to get to know each other; but I think that working together is important and the discussions of content really help that.” These data are consistent with the focus group discussion data on social cohesion, where three students (among eleven students present) described how getting to know each other was not essential, but working together to understand course content was beneficial to their learning. In other words, students viewed group work as an activity that promoted social cohesion and satisfaction, but the ability to help each other learn course material was the vital aspect for learning, not the development cohesion among group members.

With respect to student involvement, the anonymous nature of the shared whiteboard was also discussed. Four students (among twelve present) described how the anonymity of the whiteboard helped them be more involved during recitation activities, and four students expressed how it helped them learn from each other. Two of the comments from this discussion included “it helped because you can learn from yours and others mistakes while also being anonymous about it” and “Anon makes it easier for you to put an answer and contribute to the class.”

Finally, with respect to student satisfaction, when asked on the first CUCEI in August “How could recitations be improved” five out of eighteen students described that they would like to see improvements to the WC software. One such comment made in this survey included: “Wimba classroom seems to be a great learning tool, but also has many bugs.” None of the students had positive comments related to Wimba in any of the surveys, and none
of the comments in any of the surveys that related to Adobe Connect were negative. In an open-ended question that asked learners to describe technical issues in more detail, students described in the Technology Survey how issues they experienced limited their involvement in recitation activities. One example included “One day, it took me about 20-30 minutes to get Wimba to load (and by that time, a considerable part of recitation was over).” These comments suggest that technical issues impacted student involvement and satisfaction with recitation sections, and the switch from Wimba Classroom to Adobe Connect may have affected student satisfaction ratings.

**Technology use**

To characterize how students were involved in recitations via IM, the number of comments and words typed by students during recitations were determined. IM comments made during group work activities in Wimba Classroom could not be recorded, so group work data are only available from the seven activities held using Adobe Connect. The average number of comments that students wrote in whole group discussion per hour was calculated for each of the 45 recitations, and then the grand average across all 45 discussions is presented. Identical calculations were made for the group work activities (Table 4).

<table>
<thead>
<tr>
<th>Table 4. IM transcript summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole group discussion</td>
</tr>
<tr>
<td>Total number of comments</td>
</tr>
<tr>
<td>Number of sessions recorded</td>
</tr>
<tr>
<td>Average number of comments made per hour</td>
</tr>
<tr>
<td>Average words per comment</td>
</tr>
</tbody>
</table>

The overall comment rate was higher during group work activities than in whole group discussion ($t(6) = 1.65, p < .01$). The content of these comments was not examined; further research is needed before additional conclusions can be drawn from these findings. But the data in Table 4 are relevant to the research question regarding whether different learning activities affect student involvement.

A focus group discussion was facilitated at the end of the Spring 2014 semester that focused on communication during recitations with microphones. Of the thirteen participants in this discussion, five students described how having access to the whiteboard and IM was sufficient for communication during recitations. Two students did not use microphones due to the presence of others in their local physical environments that they could not disturb. Another two students described how they found microphones useful during group work.

Table 5 presents data collected from the technology survey. Responses were converted to numerical values (never = 1, rarely = 2, sometimes = 3, often = 4, always = 5) and averaged to produce the results below.

<table>
<thead>
<tr>
<th>Table 5. Technology issues encountered by students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology problem</td>
</tr>
<tr>
<td>Difficulty logging into Wimba/Adobe Connect.</td>
</tr>
<tr>
<td>Difficulty with my Internet connection.</td>
</tr>
<tr>
<td>Difficulty using text chat.</td>
</tr>
<tr>
<td>Difficulty writing on the whiteboard.</td>
</tr>
<tr>
<td>Difficulty using the Wacom tablet.</td>
</tr>
</tbody>
</table>

*Note. Never = 1, Rarely = 2, Sometimes = 3, Often = 4, Always = 5.*

Over the course of the year, the most common source of technical issues that students experienced was related to their Internet connections. Three students reported that they often experienced problems with their Internet connection in the April 2014 survey. These findings are relevant to the question of whether students could benefit from having access to web cameras, which require high connection speeds (Carrington, Kim, & Strooper, 2010). Table 5 also suggests that students had fewer difficulties logging into their WC software at the end of the year than at the beginning, and that they had more issues at the start of the year with writing on the board. Both of these findings may help explain why Satisfaction scores increased. Finally, in both September and April, students reported fewer difficulties with using text chat than with anything else, which may help explain why some students expressed during the focus group discussions that communicating via the whiteboard and IM was sufficient.
Discussion

Grade data

Differences in average final grades obtained by the high school students in the VTC and WC sections were not statistically significant. These results are encouraging: they suggest that it is possible for students to be successful in the WC delivery model that was introduced in 2012 to be a low cost alternative to the VTC model (Mayer & Hendricks, 2014). Further iterations of this delivery model with different TA's are needed to determine whether the WC format consistently produces equivalent or higher final grades than the VTC delivery format.

Involvement, cohesion, and satisfaction

Students in the WC section were given permission to write on the whiteboard anonymously at any time. During the focus group discussion on the use of the whiteboard, students expressed that the anonymous nature of this communication channel made it easier for them to be involved in recitation activities. While some educators have found that anonymous interactions in an online environment can lead to unnecessary distractions (Martin et al., 2012), others have found that anonymity in an asynchronous medium can create more authentic discussions and increase student participation (Bowen, Farmer, & Arsenault, 2012). Although to the best of the TA's knowledge, there were no instances of disruptive behavior, this could have to do with the small class size, the presence of the TA, or any number of other factors. Ultimately, the WC facilitator must be mindful of the maturity of their students when deciding the extent to which anonymous interaction should be incorporated into their sessions, and further research on the anonymous involvement is needed.

Focus group data collected in this study suggested that for some students, group work helped develop social cohesion. The small but statistically significant increase in social cohesion, as measured by the modified CUCEI, may be attributable to students getting to know each other throughout the year through group work.

The increase in satisfaction scores, as measured by the CUCEI, may have been due to the transition from Wimba Classroom to Adobe Connect. But the increase in satisfaction may have also been due to the increase in social cohesion that was observed: students described that getting to know their peers made recitations more enjoyable. Similar dynamics between student satisfaction, technical issues and social cohesion have been reported elsewhere (Kear et al., 2012; Laubach & Little, 2009; Kuo, Walker, Belland, Schroder, & Kuo, 2014). It should also be noted that for some students, the development of social cohesion was not perceived as essential. Further research on the extent to which its development is needed may require further research, particularly among advanced high school students in fully online courses.

How students used technology to communicate

Results presented in Table 4 gives us a clearer picture of how and when students used IM during recitation activities. That IM tends to be associated with fewer technical issues than other channels, and that students became more involved through IM during group work than in whole group discussion, are consistent with findings of recent studies (Bower & Hedburg, 2010, p. 475; Cornelius, 2014, p. 267). Average comment rates and lengths can be compared to those reported in the synchronous CMC literature (Lobel, Swedburg, & Neubauer, 2002; Martin et al., 2012; Oztok et al., 2013). Differences between comment rates and lengths found in this study and others could be attributed to several factors, including the nature of the activities the students were engaged in. Moreover, comment rates and comment counts cannot by themselves describe the quality of student interactions (Hrastinski, 2006, p. 140). A content analysis, such as that used by Hou and Wu (2011) or Bower and Hedburg (2010), would provide a deeper understanding of the nature of the discussion.

Study data identified challenges related to providing students access to microphones and web cameras. The video stream from web cameras requires additional Internet bandwidth, and three students in this study reported that they often experienced issues with the quality of their Internet connections. Five students wrote that they were able to communicate sufficiently with IM and the whiteboard, and two students were unable to connect in areas where they could use microphones. These findings are consistent with those reported on a recent case study that found that high school students preferred to communicate via text in WC environments (Murphy, Rodríguez-Manzanares, & Barbour, 2011, p. 590). Requiring that all students have access to web cameras and microphones necessarily increases costs for a program whose purpose is to provide a low-cost alternative to a more expensive...
delivery model. Combining these findings together, it does not seem likely that future students participating in the DCP WC delivery model will be expected to have access to web cameras and microphones.

Finally, students consistently expressed frustration with the technologies they were using across all of the surveys, so a decrease in technical issues may have led to an increase over time in student satisfaction scores as measured by the CUCEI. These findings are consistent with other case studies that suggest that learners using WC tools benefit from additional technical support (Cornelius, 2014; Martin et al., 2012; Olson & McCracken, 2015), and therefore have implications for the way that technical support and training could be improved in future offers of the WC format in the DCP.

Study limitations

This study is exploratory and its findings are unlikely to be generalizable to all contexts for several reasons. Results are based on a small population of advanced high school students. Secondly, the observed group dynamics may lie in the particular direction and structure that the teaching assistant facilitated. The activities, curriculum, and the direct questions that the assistant asked would have undoubtedly affected the group's approach to interaction and resulted in communication patterns that would have been different with other facilitators. Thirdly, the dynamics observed in the WC model are constrained by the features afforded by particular tools that were used. As new systems and technologies become available, the situation may change. Furthermore, interactions on the whiteboard and via audio were not recorded and therefore not studied. Non-observable communication would likely have occurred during recitations between students through face-to-face interactions between students attending the same school and via private messaging. Finally, comparisons between grades obtained by students in different recitation sections are complicated because the sections had different TAs, learning activities, grade weightings, and used different technologies.

Conclusions

Study results confirm findings from case studies in the recent distance education literature. Supporting evidence includes that group work activities are a way for learners to develop social cohesion, and can have a positive impact on student satisfaction. These findings are consistent with a recent case study that reported that learner-learner and learner-instructor interactions were significant predictors of student satisfaction (Kuo et al., 2014). Given that not all instructors incorporate group work activities in WC environments (Cornelius, 2014), further research on the benefits and challenges of facilitating group work activities in these settings is needed.

Study findings also clarify the role of audio and video channels in WC environments. While some participants felt that audio and video would have been beneficial for developing social cohesion, barriers to their use were identified. All of the identified restrictions are specific to the context of this study, but some of them have been reported elsewhere (Carrington, Kim, & Strooper, 2010; Murphy et al., 2011, p. 589). Educators who are deciding whether or not to require learners to have audio and video capabilities for WC sessions must consider the tradeoffs between the social benefits these channels afford and, among other factors, the technical requirements that their learners must meet in order to utilize them.

Our results also suggest other directions for future research related to the anonymous nature of the whiteboard. Research is needed clarify how anonymous interaction impacts student involvement in the WC environment and on how disruptive behavior is managed. Also, given that some students in this study found that the development of social cohesion was not essential, further research in this area may be needed, particularly among advanced high school students in online courses.

The findings presented in this article, along with those in the WC literature, are part of an ongoing effort to clarify how to best foster a student-centered environment in synchronous online learning platforms. The technologies and teaching practices that facilitators use, and how they develop and moderate learning activities, are just some of the decisions that impact the role instructors and administrators are faced with. As the technologies they use inevitably evolve, so too will our understanding of their role in distance education with ongoing research in this field.
Acknowledgements

The authors of this article would like to thank Tom Morley, George Wright, and Cher Hendricks for their encouragement, suggestions, and input on this work. The contents of this work were developed under a Race to the Top grant from the U.S. Department of Education. However, those contents do not necessarily represent the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

References


Lobel, M., Swedburg, R., & Neubauer, M. (2002). The eClassroom used as a teacher’s training laboratory to measure the impact of group facilitation on attending, participation, interaction, and involvement. The International Review of Research in Open and Distributed Learning, 3(2). doi:10.19173/irrodl.v3i2.112


Appendix: Technology Survey

Please indicate which of the following technical problems you have encountered during recitations since the beginning of this semester.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty with my Internet connection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Issues related to my computer hardware</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty moving in or out of a breakout room</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty writing on the white board</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty using text chat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty using the Wacom tablet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty with my computers operating system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty logging into Wimba</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty hearing the TA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty seeing the TA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty updating Java</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please indicate what you do when you encounter technical issues during recitations.

<table>
<thead>
<tr>
<th>Action</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>I log in and out of Wimba</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I refresh my web browser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I reboot my computer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I contact Wimba technical support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I get help from another student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I get help from a teacher or staff at my school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I get help from my teaching assistant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I resolve the issue myself</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please indicate how often you experience the following situations.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>I experience technical issues in recitations that make it difficult to participate in recitations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I experience technical issues in recitations that make it difficult to learn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I experience technical issues in recitations that make it difficult to understand what other students are saying.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please indicate how you connect to the Internet during recitations.

<table>
<thead>
<tr>
<th>Connection</th>
<th>Never</th>
<th>Rarely</th>
<th>Sometimes</th>
<th>Often</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>I connect using a wireless connection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I connect using a wired connection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I connect from home</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I connect from school</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Exploring Long-term Behavior Patterns in a Book Recommendation System for Reading

Tzu-Chao Chien1, Zhi-Hong Chen2* and Tak-Wai Chan1

1Graduate Institute of Network Learning Technology, National Central University, Taoyuan, Taiwan //
2Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taipei, Taiwan // brian@cl.ncu.edu.tw // zhchen@ntnu.edu.tw // chan@cl.ncu.edu.tw

*Corresponding author

(Submitted September 22, 2015; Revised April 20, 2016; Accepted June 7, 2016)

ABSTRACT
This study explored the behavior of students who used a book recommendation system, specifically the My-Bookstore system, over a five semester period. This study addressed two main research questions, the first being related to the most frequent behaviors and behavioral patterns. The results showed that “visiting” behavior and “book recommendation” behavior were the two most frequent behaviors in the first and last semesters. In addition, there was a tendency for the recommendation behavior to increase over the five semesters, implying that the My-Bookstore system can facilitate the book recommendation behavior. The second research question concerned the influences of incentive models on the recommendation behavior. The findings demonstrated that the coin reward incentive in the game has a short-term impact on student recommendation behavior. However, those functions that have an internally tightly-coupled relevance to book recommendation (e.g., “choosing recommended books” as a store management incentive, and “checking histories” a social interaction incentive) act to give the My-Bookstore program better long-term effects on the students’ book recommendation.

Keywords
Book recommendation system, Incentive models, Behavior analysis

Introduction
Book recommendation between students has the potential to enhance reading interest and ability (Chien, Chen, Ko, Ku, & Chan, 2015). The recommendation process encourages students to share and discuss what they have read, thus helping to create and foster a good reading atmosphere within the class (Larson, 2009). A book recommendation system can provide students with an opportunity to review what they have read by retelling, based on their own individual understanding and perspectives, which can greatly improve their literacy skills, reading comprehension and recollection (Carico, Logan, & Labbo, 2004). However, the benefits of a book recommendation system are reliant upon extensive and frequent student interaction, which requires overcoming the difficulties of organizational complexity and limited interaction. The former relates to the problem of class time allocation. It is difficult for teachers in the classroom to clear space on the agenda for a book recommendation activity and ensure that all students have an equal chance to express their opinions. The latter refers to the time needed in order to reinforce the effect of social interaction. In small class size elementary classrooms, there might not be enough student interaction for peer-to-peer book recommendation.

Strategies proposed to overcome these two difficulties include activities designed to reduce the complexity and enhance the efficiency of social interaction, such as book talks, storytelling, and dramas (Pilgreen, 2000; Gardiner, 2005; Atwell, 2007). Although these activities can greatly help to facilitate classroom book recommendation, they still require the teacher to expend a lot of effort on logistical organization, such as setting up tables, group allocation, and maintaining classroom order. Recent developments in information technology however, have opened the door for teachers to provide tech-support support for the development of communication tools and book recommendation systems (Hamilton & Cherniavsky, 2006, Larson, 2008) in the classroom, include message board discussions (Wolsey, Biesenbach-Lucas, & Meloni, 2004), blogs (Huffaker, 2005; Ray, 2006), and social networking systems. With the support of technology, teachers can efficiently plan activities to help students express their opinions to their peers, both in and out of school time (Hancock, 2008). These technology-supported activities can serve as a stage, on which students can express ideas, perspectives, and thoughts to their classmates, and help to foster the development of a learning community for reading and book recommendation (Wolsey, Biesenbach-Lucas, & Meloni, 2004; Larson, 2009).

One example is to use potential of technology for promoting the behavior of book recommendation. The My-Bookstore system (Chien, Chen, Ko, Ku, & Chan, 2011), which incorporates incentive models into a recommendation system, is designed to help students describe their favorite books and recommend them to their classmates. Although a previous study has demonstrated that such a system can enhance student learning in...
terms of word usage and reading perception (Chien et al., 2015), little attention was paid to investigating the students’ behavior patterns over a long period of time. Investigating student behavior when using such a book recommendation system is critical because not only do the findings enrich our understanding of how an effective system works, but also sheds light on how to stimulate and maintain positive behaviors. The purpose of this study is thus to conduct a long-term empirical study using the My-Bookstore system as an example. In addition, to acquire a more comprehensive understanding of student behavior, we not only analyze the frequency of certain behaviors, but also trace their sequence. In this way, subtle changes in behavior can be revealed, even if the frequencies remain similar. The two research questions to be answered in this study are: (1) What are the frequent behaviors and behavioral patterns of students participating in a book recommendation system? (2) What incentive models for such a book recommendation system can significantly facilitate students’ recommendation behavior?

Book recommendation system - My-Bookstore

Design principles

My-Bookstore is a book recommendation system designed to encourage book recommendation among elementary students in a classroom environment (Chien et al., 2011). The development of My-Bookstore is underpinned by two design principles, the “open student model” (Bull & Kay, 2007) and a theory “learning community” (Bruckman, 2006). Based on the open student model students are helped to understand what they have learned with the support of technology, leading to enhanced self-awareness, self-reflection, and self-improvement (Chen, Chou, Deng, & Chan, 2007; Bull, Gardner, Ahmad, Ting, & Clarke, 2009; Chen, 2012). This “open” model is beneficial to students (Bull, 2004; Vélez, Fabregat, Bull, & Hueva, 2009). The visualization of a collection of student work (or portfolios) serves as a window to understand their learning effort, progress, and achievements (Paulson, Paulson, & Meyer, 1991; Chang & Chen, 1998).

In contrast, the learning community refers to a group of learners who share common goals and attitudes, and accomplish learning tasks by communication, discussion, and collaboration (Gabelnick, MacGregor, Matthews, & Smith, 1990; Bruckman, 2006). In a sense, a classroom is a small social arena, where a student’s concepts, beliefs, and behavior are affected by their classmates (Alsop & Hicks, 2013). For instance, if a student finds that reading is interesting and meaningful, his/her behavior and attitude might be imitated by the other students, subsequently spreading to the whole class. Thus, in the “My Bookstore” system, each student runs their own “bookstore”, which, when taken altogether become a bookstore “community,” helping to create a positive reading atmosphere.

These two design principles have guided the development of the My-Bookstore system. In the My-Bookstore system each student acts as a manager running his/her own bookstore, and can recommend the books that he/she has read to other students. In the My-Bookstore system, “stocking” of the bookstore is done by means of students recording the books they have read, and “selling” means recommending the books they like to others. In other words, the books in the My-Bookstore system are a concrete representation of the student’s reading profile, a record of what each student has read, and a useful medium to recommend books to their peers.

Book recommendation

In the system, there are four ways to make a recommendation: star-ranking, picture-drawing, writing, and sound-recording (see Figure 1). The recommendation process can be divided into three steps: (1) the students first read a book borrowed from the classroom library, which when record in the system, will automatically show up in their historical bookcase; (2) students can comment about the books they have read by means of the four recommendation choices mentioned above; (3) student can monitor which of the books they have recommended are accepted by their peers.

It was found in a previous study that most of the students using the My-Bookstore system were willing to recommend their favourite books (Chien et al., 2015). Furthermore, of the four recommendation choices, the most popular was the start-ranking function (96% of students used this function), because its simplicity made it easy for even young students to complete the task. The second most popular was picture-drawing (43%) because drawing is a natural way for children to express ideas. The least two popular functions were writing (25%) and sound-recording (19%), probably because students lacked sufficient experience in writing, recording, and self-expression.
Three incentive models

Three types of incentives (see Figure 1) were developed for application in the My-Bookstore system: virtual coin rewards, store management, and social interaction. The coin reward incentive allows students to earn virtual coins (i.e., called Book-Coins) as a reward for enthusiasm and effort. Book-Coins are earned by successfully recommending books to their classmates and can be used by the students to purchase new decorations for their bookstores.

The store management incentive allows students to create their own unique bookstores. They can decide how to alter the appearance of their bookstore and how to decorate them. In addition, they can also write “slogans”, phrases to be spoken by a virtual worker when someone enters the bookstore, for example “Welcome to my bookstore!” As a bookstore manager, the student has to take responsibility for the marketing income, and establish a plan for book recommendation, because each bookstore has limited bookshelves for book recommendations. In other words, they need to plan and prepare the best choices for book recommendation.

The social interaction incentive is designed to offer opportunities for students to recommend books to each other. For example, the “visiting” function allows students to visit their classmates’ bookstores. This provides students with the opportunity to become aware of their peer’s reading status, for example by seeing what books their classmates have read and their opinion of these books. Another function related to social interaction is a checklist of interaction histories. A student can check the list to find out how many people have visited their bookstore, and who has accepted their book recommendations.
Methodology

Participants

A total of 204 first-grade (aged 7-8 on average) elementary students at a suburban elementary school in Taiwan participated in this study. This school was selected because approximately one-third of the students were from families with a low or middle level socioeconomic status. The lessons learned from this study can be applied to other schools with a similarly diverse student populations in the future. The students participated in this study for two and a half years (from February 2011 to June 2013, a total of five semesters). In the first year, the students were divided into seven classes (with seven different teachers). In the second year, with the exception of a few who had transferred to other schools, students were divided into eight classes (with eight different teachers) based on the school’s management policy.

Procedure

In order to facilitate access to books, which is an important element of a promotional reading activity (Pilgreen, 2000), a small library was set up in every class. Each class library had at least two hundred books. The books were selected by participating teachers as suitable for the students. In addition, the participants used the system in a one-to-one environment (Chan et al., 2006), with every student given access to a touch-screen tablet computer so that they could use the My-Bookstore system. They were allowed to take these devices home. Every classroom had internet connection capability for system access.

The participants (both students and teachers) attended a reading activity in the morning - the Modeled Silent Sustained Reading program (McCracken, 1971; Pilgreen, 2000; Gardiner, 2005; Chien et al., 2011), where they were allowed to read extracurricular paper books for 20 to 30 minutes, four days a week (see Figure 2). After the reading activity, students were free to use the My-Bookstore system without any requirements as to the number of books they must read or recommend.

Data collection

Data were collected and student behavior was logged within the My-Bookstore system using the User-Behavior-Time format, where User indicates the name of the student, Behavior refers to the specific functions used in the My-Bookstore system (see Table 1), and Time is the time stamp showing date: hour: minute: second. Data analysis was then conducted based on the system logs.

Data analysis

Two methods of data analysis were employed, based on the aforementioned system logs: frequency analysis and sequential analysis (Jeong, 2005; Bakeman & Gottman, 1997). First, frequency analysis was carried out to find the frequency and percentage of each behavior demonstrated by all students, so as to present an overall picture of system use. Second, sequential analysis was conducted to emphasize the patterns of behavior based on the time sequence, in an attempt to gain insights leading to an understanding of the relationship between these behaviors (Chen, 2014; Hou, 2012). This approach was adopted because it not only focuses on the characteristics of
particular behaviors, but also on the characteristics of the interactions or transactions among those behaviors over time. This approach has been used in early works to investigate student behavior patterns, and it offers insights into understanding the relationship between these behaviors (Jeong, 2005; Bakeman & Gottman, 1997).

Table 1. Description of the specific behaviors

<table>
<thead>
<tr>
<th>Incentive</th>
<th>Behavior</th>
<th>Description</th>
</tr>
</thead>
</table>
| Book Recommendation| Making book recommendations (MR) | • Use the star-ranking (one to three) to recommend a book, and click on the appropriate reasons.  
• Draw a picture related to the book read with the drawing function of My-Bookstore.  
• Write down your favorite sentence or paragraph from the book read; give the reason you like this book.  
• Voice the opinions about or read aloud from your favorite paragraph with the sound recording function of My-Bookstore. |
| Social Interaction | Visiting a peer’s bookstore (VB) | • Click on the portrait button in the galaxy system to view a classmate’s bookstore in My-Bookstore. |
|                    | Checking histories (CH)          | • Use the historical interaction checklist interface to find out which classmate has accepted your book recommendations. |
| Store Management   | Decorating the bookstore (DS)    | • Arrange new decorations or move existing ones to another place.             |
|                    | Choosing recommended books (RB)  | • A new recommendation can be added to the bookshelf so that the contents can be publicly viewed by visiting classmates.  
• A previous recommendation can be removed from the shelves. |
|                    | Setting a slogan (SS)            | • Certain sentences can be set as messages for the virtual character to speak when someone comes to the bookstore. |
| Virtual Coins reward| Getting rewards (GR)            | • Use “Book-Coins” to purchase new bookstore decorations from the store. |

Results and discussion

Behavior frequency analysis

Table 2 shows the frequency and percentage of different behavior demonstrated over the five semesters. Overall, the behavior of "visiting a peer’s bookstore (VB)" was the most frequent over all of the semesters (52% to 34%), followed by “making book recommendations (MR)” (11% to 29%), “decorating one’s personal book store (DS)” (11% to 14%), and “getting rewards (GR)” (8% to 11%). The results also showed an increased trend in the MR behavior over the five semesters, implying that the target behavior (i.e., MR) was successfully promoted.

Table 2. The frequency and percentage of behavior-codes over the semesters

<table>
<thead>
<tr>
<th>Semester</th>
<th>Book Recommendation</th>
<th>Social Interaction</th>
<th>Store Management</th>
<th>Virtual Coin Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MR</td>
<td>VB</td>
<td>CH</td>
<td>SS</td>
</tr>
<tr>
<td>1st</td>
<td>freq.</td>
<td>6275</td>
<td>18106</td>
<td>1708</td>
</tr>
<tr>
<td></td>
<td>(pct.)</td>
<td>(18%)</td>
<td>(52%)</td>
<td>(5%)</td>
</tr>
<tr>
<td>2nd</td>
<td>freq.</td>
<td>9221</td>
<td>19660</td>
<td>2342</td>
</tr>
<tr>
<td></td>
<td>(pct.)</td>
<td>(23%)</td>
<td>(48%)</td>
<td>(6%)</td>
</tr>
<tr>
<td>3rd</td>
<td>freq.</td>
<td>9117</td>
<td>20622</td>
<td>2430</td>
</tr>
<tr>
<td></td>
<td>(pct.)</td>
<td>(20%)</td>
<td>(46%)</td>
<td>(5%)</td>
</tr>
<tr>
<td>4th</td>
<td>freq.</td>
<td>8946</td>
<td>11406</td>
<td>1889</td>
</tr>
<tr>
<td></td>
<td>(pct.)</td>
<td>(28%)</td>
<td>(36%)</td>
<td>(6%)</td>
</tr>
<tr>
<td>5th</td>
<td>freq.</td>
<td>5411</td>
<td>6389</td>
<td>979</td>
</tr>
<tr>
<td></td>
<td>(pct.)</td>
<td>(29%)</td>
<td>(34%)</td>
<td>(5%)</td>
</tr>
</tbody>
</table>
In addition, the results for behavior frequency pointed out that the two most frequent user behavior were “visiting other’s bookstores” (VB) and “making book recommendations” (MR). The former showed a decreasing trend, whereas the latter demonstrated an increasing trend. Although the VB behavior showed a tendency to decrease, it was still the most common behavior each semester. In addition, the decrease of VB is almost transformed into the proportional increase of the target behavior (i.e., MR). This finding seems to suggest that initially, social interaction incentives played the role of “motivator” to sustain students’ participation, but this gradually transformed into the role of “facilitator”, promoting the target behavior. The lack of change in the frequency is not equivalent to a lack of change in behavior, because there are subtle differences in behavioral sequences. Thus, there is a need to examine the results with behavioral sequential analysis.

**Behavioral sequential analysis**

The behavioral patterns for the first and fifth semesters obtained through the sequential analysis matrix calculations are illustrated in Figure 3 and Figure 4, respectively.

**Figure 3.** Sequential status of the seven behaviors in the first semester

**Figure 4.** Sequential status of the seven behaviors in the fifth semester
Virtual coin reward incentive

As shown in Figure 3, the behavior pattern goes from “making book recommendation” (MR) to “getting rewards” (GR) in the first semester. In addition, the patterns of “getting rewards” (GR) → “decorating the store” (DS), and “decorating the bookstore” (DS) → “getting rewards” (GR) were also significant. It seems that the incentives supplied by the virtual coin reward and decoration of the bookstore initially attracted the students’ attention. The students who received the coin rewards tended to go to the store to purchase decorations. However, the pattern of MR→GR in the fifth semester was not significant, as shown in Figure 4, implying that the virtual coin rewards did not play a critical role for book recommendation over a long period of time.

A possible explanation was that the coin reward incentive is an external representation of how much effort the student has exerted. Although this approach is frequently used in game-based systems (Zichermann & Cunningham, 2011; Nicholson, 2012; Nicholson, 2013), the results demonstrate its limitations. The coin reward incentive might be suitable in the beginning, but is inappropriate over a long period of time. This finding can be interpreted based on the theory of interest development (Hidi, 2006), which argues that interest (e.g., in this case the reading and sharing of good books) can be developed starting from situational interest moving to individual interest (Hidi & Renninger, 2006). Specifically, individual interest involves personal preferences and characteristics. It develops slowly, but has long-lasting effects. Inversely, situational interest is evoked suddenly by certain stimuli in the environment, but has only a short-term effect (Hidi, 1990). Thus, for those students who do not have a personal interest in reading and book sharing in the beginning, situational interest (i.e., the rewards approach) can serve as a stimulus to cultivate the target behavior. The findings suggest that virtual coin rewards can be an effective incentive (triggered by the situation) in the beginning, but its influence will not last for a long time. This should be taken into account by designers who can take the opportunity to shape student behavior by first stimulating through situational interest and then by nurturing individual interest through well-developed guidelines for learning systems.

Store management incentive

Regarding the store management incentive, in the first semester, “decorating the bookstore” (DS) did not significantly contribute to the “making recommendation” (MR) behavior. However, the “setting a slogan” (SS) and “choosing recommended books” (RB) behaviors were followed by the book recommendation behavior (i.e., SS→MR and RB→MR), implying that both SS and RB were important components in the beginning. However, the results in the fifth semester told a different story: the pattern of SS→MR was no longer significant, implying that SS has a limited long-term impact on book recommendations. Rather, the major component of store management that really contributes to book recommendation is RB.

Of the three behavior patterns related to the bookstore management incentive (i.e., DS, SS, and RB), only RB had a long-lasting impact on students. From the perspective of the open student model (Epp, & Bull, 2015; Bull et al., 2009), the major difference between the three behavior is the level of coupling between the “open student model” and “what the students do.” The bookstore is thus a visualization of the student model, so taking good care of the bookstore is actually taking good care of their learning. The results reveal that RB is more effective in terms of instilling active and responsible management behaviors with a tightly-coupled relationship with the student model (i.e., books they read). In contrast, the other two behaviors (i.e., DS and SS) focus on increasing the attractiveness of the environment, and are only loosely-coupled to management. Specifically, neither “setting a slogan” (SS) nor “decorating the bookstore” (DS) are really relevant to book recommendation, and neither serve to maintain student behavior in the long run. This finding suggests that the management incentive can be regarded as self-regulation of the open student model. Thus, management incentives can be categorized as tightly-coupled or loosely-coupled, depending on the relevance of their interaction and regulation with the open student model. Designers should increase the coupling level of management incentives with the open student model (in this case it is the books that they have read).

Social interaction incentive

Regarding the social interaction incentive, in the first semester, the results showed that the pattern of “visiting a peer’s bookstore” (VB) to “making a book recommendation” (MR) was not significant, but the pattern of “checking histories” (CH) to “making a book recommendation” (MR) was significant. The fifth semester results were the same. It can thus be stated that the major component of social interaction as it contributes to book
recommendation lies in CH (i.e., CH→MR). Students were interested in the historical interaction checklist, which they used to understand which classmate had accepted their book recommendations.

From the findings of behavioral sequential analysis, the “visiting a peer’s bookstore” (VB) did not seem to play a critical role in promoting the target behavior. However, the results of behavior frequency indicate that VB could play a “motivator” role in sustaining students’ participation. The role of VB can be confirmed as maintaining continuous participation, even though it did not contribute to the promotion of the target behavior. On the other hand, the behavior of “checking histories” (CH), which also belongs to the category of social interaction, did contribute to the promotion of the target behavior. From the perspective of social interaction, the major differences between VB and CH are the differences in approaches to facilitating self-reflection in the learning community: VB involves the self-reflection via the outside-looking approach (i.e., observing peers), whereas CG involves the self-reflection via an inside-looking approach (i.e., observing comments made by peers). In other words, the looking-inside approach can directly foster self-reflection. Thus, the findings seem to suggest that the use of the social interaction incentive would be an appropriate approach for fostering self-reflection.

Conclusions

This study addressed two research questions. The results related to the first (i.e., What are the most frequent behaviors and behavioral patterns of students participating in a book recommendation system?) showed “visiting other people’s bookstores” (VB) and “making book recommendations” (MR) were the two most frequent behaviors in the first and last semesters. The results also revealed an increasing trend in the MR behavior over the five semesters, implying that the My-Bookstore system can facilitate the book recommendation behavior.

Regarding the second research question (i.e., What incentive models for such a book recommendation system can significantly facilitate students’ recommendation behavior?), the findings demonstrated that the coin reward incentive has a short-term impact on student recommendation behavior. However, functions that have an internally tightly-coupled relevance to book recommendation (e.g., “choosing recommended books” (RB) in the store management incentive, and “checking histories” (CH) in the social interaction incentive) act to give the My-Bookstore program better long-term effects on the students’ book recommendation.

This system was developed to encourage book recommendation behaviors through the incorporation of three incentive models, including virtual coin rewards, store management, and social interaction. A long-term empirical study over two and a half years was conducted to determine what the most frequent behaviors and behavioral patterns of students participating in the book recommendation system are. In addition, the issue as to which incentive models could most significantly facilitate students’ recommendation behaviors was also discussed. The results indicated that: (1) My-Bookstore can facilitate student engagement in peer interaction and book recommendations; (2) the coin reward incentive has only a short-term impact on book recommendation behavior and the store management incentive and social interaction incentive have better long-term effects.

There are some limitations to this study that should be further investigated in the future. First, although this study has shown the patterns of students’ behavior in using the My-Bookstore system, a systematic design of control groups is lacking. Second, quality analysis does not provide a rich picture of the complicated status of system usage, which is why we need qualitative analysis to further clarify the system impact on students’ recommendation behaviors.

Acknowledgments

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan, for their financial support (104-2511-S-008-009-MY3), and the Research Center for Science and Technology for Learning, National Central University, Taiwan.

References


Learning through Blogging: Students’ Perspectives in Collaborative Blog-Enhanced Learning Communities

Yu-Chun Kuo1*, Brian R. Belland2 and Yu-Tung Kuo3

1Department of Science, Technology, Engineering, Art and Math Education, Rowan University, Glassboro, NJ, USA // 2Department of Instructional Technology and Learning Sciences, Utah State University, Logan, UT, USA // 3Department of Computer Graphics Technology, Purdue University, West Lafayette, IN, USA

yu.chun100@gmail.com; kuo@rowan.edu // brian.belland@usu.edu // ytkuo1005@gmail.com

*Corresponding author

(Submitted September 25, 2015; Revised February 07, 2016; Accepted July 10, 2016)

ABSTRACT

This study employed a mixed method approach to investigate the relationships between learners’ blogging self-efficacy, sense of community, perceived collaborative learning, and perceived learning in classroom environments. Learners’ perspectives of group learning experiences in blog-enhanced settings were examined. Participants were minority adult students enrolled in two courses offered at a university in the southern United States. Results indicated that (a) sense of community and perceived collaborative learning significantly contributed to perceived learning through blogging; (b) blogging self-efficacy was not a good predictor of perceived learning but was related to prior experiences of using blogs; (c) most students displayed positive attitudes toward the use of blogs and group learning experiences that involved collaborative process as well as the development of knowledge and skills; and (d) individual dispositions had a potential influence on collaboration. This study adds to the limited research on the use of blogging and collaborative work among minority students in continuing education. Results are discussed in light of the literature and suggestions for future research are provided.

Keywords

Collaborative learning, Sense of community, Blogging self-efficacy, Mixed methods, Constructivism

Introduction

Blogs have been increasingly utilized in higher education to facilitate student learning (Halic, Lee, Paulus, & Spence, 2010; Shana & Abulibdeh, 2015; Top, 2012). They allow users with no advanced programming skills to create an online space through posting, editing, and publishing articles composed of text, images, audio, video, and hyperlinks (Papastergiou, Gerodimos, & Antoniou, 2011). In blog-supported environments, learners learn to and begin to appreciate information sharing, idea exchange, and collaboration, which in turn contributes to student learning or professional development (Top, 2012; Wassell & Crouch, 2008). Research has indicated that blogs can serve as effective teaching and learning tools to (a) support students’ active participation through collaboration in a class (Kiliç, & Gökdaş, 2014; Top, 2012), (b) enhance peer support and interaction (Chang & Chang, 2014; Laal & Laal, 2012), (c) increase students’ motivation to learn the subject content (Pursel & Xie, 2014; Shana & Abulibdeh, 2015), and (d) develop students’ critical thinking and reflective skills (Li, Bado, Smith, & Moore, 2013; Ellison & Wu, 2008; Xie, Ke, & Sharma, 2008).

The majority of studies on blogging in education have been completed in higher education (undergraduate- and graduate-level) and K-12 settings across various disciplines (Shana & Abulibdeh, 2015; Top, 2012; Xie, Ke, & Sharma, 2008), and limited focus has been placed upon non-traditional adult students in continuing education (Park, Helo, & Lee, 2011). Blogs may help create a meaningful learning environment for adult students, and to some degree, blogs may also cater to adult students’ learning preferences, including providing for self-direction, practical experiences, and student-centered learning (Merriam, Caffarella, & Baumgartner, 2007). Blogs can facilitate positive learning experiences among adult learners through reflection-oriented learning processes (Park, Helo, & Lee, 2011). There is little research focusing on minority students’ perceptions of learning experiences with blogs. This study involves African American students in continuing education. African-American students are considered as high context learners who have a preference to work in groups rather than work independently or individually (White, 1992). Participating in a learning community appears to increase the chance of academic success among African-American students (Duncan & Barber-Freeman, 2008; Gallien & Peterson, 2004). It would be valuable to further explore African-American adult students’ learning experiences with blogs.
In addition, although previous research indicated that the use of blogs enhances students’ collaborative learning experiences and sense of community, little research describes effective strategies to integrate blogs into instruction so as to enhance knowledge acquisition, collaboration and communication skills, belongingness, and affective outcomes in classroom settings (Top, 2012). Presumably, one’s confidence in executing blogging-relevant tasks is important and relevant to an individual’s learning experience with blogs. However, blogging self-efficacy was not included in any previous research studies as a predictor of students’ perceived learning in blog-enhanced settings. Furthermore, no previous study has investigated blogging self-efficacy, collaborative learning, and sense of community together, as well as the effect of these three factors on students’ perceived learning of blogging among African American adult students. Therefore, the purpose of this study was to investigate the relationships between blogging self-efficacy, sense of community, collaborative learning, and perceived learning in blog-enhanced learning settings.

Theoretical framework

According to the social constructivism embodied by cultural-historical activity theory, knowledge is constructed through social interactions, including conversation, discussion, and negotiation processes (Leont’ev, 1974; Luria, 1976; Rondon-Pari, 2011; Vygotsky, 1978). The zone of proximal development (ZPD), intersubjectivity, mediation, and enculturation are four concepts underlying the process of learning according to social constructivist perspectives (Engeström, 2001; Woo & Reeves, 2007). ZPD is closely tied to assessment, in that it posits that the focus of assessment should be on students’ problem-solving abilities when collaborating with more capable peers or adults, rather than on students’ independent problem solving abilities (Vygotsky, 1978; Wertsch, 1984). The ZPD is then the distance between an individual’s problem-solving abilities when assessed in these two different ways. Intersubjectivity refers to a shared understanding of the goals of instruction (Verenikina, 2008; Woo & Reeves, 2007). According to enculturation, students learn knowledge, skills, and values through authentic engagement in a target culture (Kottak, 2007). The use of mediating artifacts (e.g., tools, signs, symbols, language) is relevant to the interaction of subjects (e.g., actors) and objects (e.g., goals; Engeström, 2001; Hogan & Tudge, 1999). That is, individuals’ learning is shaped through tool mediation and changes of cultures over time.

Learning happens through meaning making that involves the process of sharing various perspectives and experiences in communities of practice (Vygotsky, 1978). There are different levels of collaboration involved during meaning negotiation where learners interact with peers to construct information together in social and cultural contexts (Cuhadar & Kuzu, 2010; Sivan, 1986). Collaborative groupwork (e.g., group discussions, team projects) enhances meaning construction and negotiations from multiple perspectives (Smith & Ragan, 2005; Woo & Reeves, 2007; Zhu, Valcke & Schellens, 2009). Using technological tools, such as blogs, in constructivist learning environments provides learners the means/tools to cooperate and interact with peers from different cultures, seek and share information, and solve problems and make decisions (Cuhadar & Kuzu, 2010).

Collaborative learning

In collaborative learning, shared learning experiences allow learners to engage in discussion, converse with other learners, and present or defend ideas, which enhances not only interaction among learners, but also critical thinking and problem-solving skills (Armstrong & Hyslop-Margison, 2006; Kuo & Belland, 2016; Kuo, Walker, Belland & Schroder, 2013; Gokhale, 1995; Smith & MacGregor, 1992). At its core, collaborative learning is learner-centered. But it additionally involves multiple students working together to accomplish common goals. Collaboration can be facilitated with various forms of communication in face-to-face or computer-supported settings (Laal & Laal, 2012).

Blogs support collaborative learning by enhancing knowledge acquisition, knowledge sharing, and reflective processes (Wang, 2010; Yang & Chang, 2012). For example, Wang (2010) investigated students’ perceptions of utilizing blogs as a platform for content review, data collection, and idea sharing in collaborative groups. Students perceived that blogs (a) are a useful tool to reflect and interact with classmates, and (b) enlarged the resources of learning support. Yang and Chang (2012) examined the influence of integrating blogs as supplementary tools on student learning in a traditional instructor-led class. Blogging was found to be a medium that enhances asynchronous peer interaction, reflection, and positive attitudes toward academic achievement in collaborative activities.
Sense of community

Sense of community refers to “a feeling that members have of belonging, a feeling that members matter to one another and to the group, and a shared faith that members’ needs will be met through their commitment to be together” (McMillan & Chavis, 1986, p. 9). Sense of community can be determined through shared goals and responsibility, student–instructor interaction, learner-learner interaction, value and interest, peer respect, and emotional connection (Cho, Bang, Mathew, Bridges, & Watson, 2010; McMillan & Chavis, 1986). Students’ sense of community is positively related to learning outcomes, including learning experiences, achievement, and student effort (Sadera, Robertson, Song, & Midon, 2009; Sánchez, Colón & Esparza, 2005). Furthermore, from a self-determination theory perspective, sense of community is similar to the concept of belongingness, which is one of the three motivational needs (i.e., competence, relatedness, and autonomy; Ryan & Deci, 2000).

Blogs can enhance sense of community by increasing participation, engagement, and interaction in classroom or online learning (Cuhadar & Kuzu, 2010; Yang, 2009). Learners are more engaged when in blog-enhanced settings than when in other web-based environments (Yang, 2009). Appropriate use of blogs with instructional strategies that align with course content contributes to feelings of a learning community. Through peer dialogue, learners develop reflective thinking and self-examination skills that result in personal growth and knowledge acquisition (Cuhadar & Kuzu, 2010; Xie, Ke, & Sharma, 2010). Blogs can enhance interpersonal communication skills and group interactions, especially when students are asked to provide feedback or comments (Blau, Mor, & Neuthal, 2009).

Blogging self-efficacy

Blogging self-efficacy refers to confidence in performing multimedia tasks in blogging, including capturing multimedia, multimedia processing, and content transfer to a blog. Research has shown that blogs can be a useful tool for formal and informal learning in various disciplines (Park, Heo, & Lee, 2011). Students’ confidence in using blogs may influence their learning experience in a learning environment where blogs are required (Top, 2012). Blogging self-efficacy is associated with one’s self-efficacy in using information and communication technologies (Papastergiou, Gerodimos, & Antoniou, 2011). Research on the impact of blogging self-efficacy on perceived learning is scarce. Hence, we assumed that blogging self-efficacy may influence perceived learning in this study.

Perceived learning

Much research on the effectiveness of blogging for instruction focuses on its impact on objective measures of learning (e.g., grades, quality ratings of student-produced artifacts; Papastergiou et al., 2011; Shobieri, Rashidi, Meiboudi, & Saradipour, 2014; Xie, Ke & Sharma, 2008; Xie, Ke & Sharma, 2010). While such research is important, it is also important to measure students’ perceptions of learning. First, research among college students shows that perceived usefulness of learned material is highly correlated with perceived learning (Roszkowski & Soven, 2010). When students perceive that they learned a lot, they may also be more likely to perceive that the content that they learned was useful, and in turn to apply the content when applicable. Second, perceived usefulness is critical to motivation in that it informs the value that students assign to the content to be learned (Brophy, 1999; Wigfield & Eccles, 2000). Together with expectancy for success, perceived value drives student motivation (Wigfield & Eccles, 2000), and perceived value can be particularly important in driving lifelong learning (McCombs, 1991).

Research questions

- What are learners’ blogging self-efficacy, sense of community, perceived collaborative learning, and perceived learning?
- What is the relationship between learners’ blogging self-efficacy, sense of community, perceived collaborative learning, and perceived learning?
- Do learners’ blogging self-efficacy, sense of community, and perceived collaborative learning predict perceived learning?
- Do learners who were aware of blogs before attending the class show better blogging self-efficacy and perceived learning than those who were not?
- What are learners’ perspectives of group learning experiences through the use of blogs?
Method

Sample

Participants were African American undergraduate students enrolled in two instructional design courses ($N = 27$ and 33, respectively) from a southern university in the United States (See Table 1). The courses were taught in the fall and spring semester, respectively, in the same academic year. The students were non-traditional adults who often have a full-time job and take evening or online courses for degree-seeking purposes in continuing education. The response rate was 89.5%. Most students (61.7%) were not familiar with blogs prior to starting the class. Among those who were familiar with blogs before the class, 52.2% had used blogs.

<table>
<thead>
<tr>
<th>Table 1. Background information of participants</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>10</td>
<td>16.7%</td>
</tr>
<tr>
<td>Female</td>
<td>50</td>
<td>83.3%</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>15</td>
<td>25%</td>
</tr>
<tr>
<td>26-35</td>
<td>8</td>
<td>13.3%</td>
</tr>
<tr>
<td>36-45</td>
<td>18</td>
<td>30%</td>
</tr>
<tr>
<td>46-55</td>
<td>11</td>
<td>18.3%</td>
</tr>
<tr>
<td>Above 56</td>
<td>8</td>
<td>13.3%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>60</td>
<td>100%</td>
</tr>
<tr>
<td>Caucasian</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Asian</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Knew about blogs before taking the course</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>23</td>
<td>38.3%</td>
</tr>
<tr>
<td>No</td>
<td>37</td>
<td>61.7%</td>
</tr>
<tr>
<td>Used blogs before taking the course (Only for those who knew about blogs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>12</td>
<td>52.2%</td>
</tr>
<tr>
<td>No</td>
<td>11</td>
<td>47.8%</td>
</tr>
</tbody>
</table>

Data collection

Printed surveys were distributed at the end of the group project. To increase response rate, participants were awarded extra points. Students in the spring course were required to respond to one open-ended question that asked students about their perceived learning experiences of groupwork through blogging.

The survey included five subscales covering learner background, blogging self-efficacy, sense of community, perception of collaborative learning, and perceived learning (see Table 2). The latter four subscales were used previously and high reliability values were found, except for blogging self-efficacy whose original reliability was not reported by the developers (see Table 2). The validity of scores resulting from administration of these four scales was examined by the original developers using factor analysis.

<table>
<thead>
<tr>
<th>Table 2. Instruments</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scales</td>
<td>Scale type</td>
<td>Number of items</td>
</tr>
<tr>
<td>Blogging self-efficacy</td>
<td>5-point Likert</td>
<td>13</td>
</tr>
<tr>
<td>Sense of community</td>
<td>5-point Likert</td>
<td>6</td>
</tr>
<tr>
<td>Perceived collaborative learning</td>
<td>5-point Likert</td>
<td>8</td>
</tr>
<tr>
<td>Perceived learning</td>
<td>5-point Likert</td>
<td>7</td>
</tr>
</tbody>
</table>

Note. NA refers to “not available” as the reliability information of blogging self-efficacy was not reported in the original study.

The blogging self-efficacy scale measured students’ perceptions of their ability to perform tasks relevant to multimedia blogging, including multimedia capturing, processing, and blogging (Papastergiou, Gerodimos, & Antoniou, 2011). The collaborative learning scale measured learners’ preferences of group versus individual work, preferences of online versus face-to-face interaction, perceptions of collaboration, and overall satisfaction.
with collaborative learning (So & Brush, 2008). The sense of community and perceived learning scales contained 6 and 7 items, respectively (Halic, Lee, Paulus, & Spence, 2010). The sense of community scale measured students’ perceptions of the extent to which using blogs facilitated community building. The perceived learning scale measured the extent to which students perceived that using blogs enhanced their learning. Table 3 provides a list of items in each scale.

### Table 3. Items of each scale

<table>
<thead>
<tr>
<th>Scales</th>
<th>Items</th>
</tr>
</thead>
</table>
| **Blogging self-efficacy** | 1. Transferring photos from a mobile phone to a computer  
                          | 2. Using a mobile phone to shoot video                              |
|                         | 3. Using a digital camcorder to shoot video                          |
|                         | 4. Using a mobile phone to take photos                              |
|                         | 5. Transferring video from a mobile phone to a computer              |
|                         | 6. Transferring photos from a digital camera to a computer           |
|                         | 7. Using a digital camera to take photos                            |
|                         | 8. Transferring video from a digital camcorder to a computer         |
|                         | 9. Using a blog to read multimedia content                           |
|                         | 10. Using a blog to comment multimedia content                       |
|                         | 11. Using a blog to publish multimedia content                       |
|                         | 12. Using video processing software                                 |
|                         | 13. Using image processing software                                 |
| **Collaborative learning** | 1. Collaborative learning experience in the computer-mediated communication environment is better than in a face-to-face learning environment.  
                      | 2. I felt part of a learning community in my group.                 |
|                         | 3. I actively exchanged my ideas with group members.                |
|                         | 4. I was able to develop new skills and knowledge from other members in my group.  
                      | 5. I was able to develop problem solving skills through peer collaboration.  
                      | 6. Collaborative learning in my group was effective.                |
|                         | 7. Collaborative learning in my group was time-consuming.            |
|                         | 8. Overall, I am satisfied with my collaborative learning experience in this course. |
| **Sense of community** | 1. I visit our nutrition blog more than required by my instructor.  
                      | 2. The blog helps me feel connected to other students in this course. |
|                         | 3. Due to the class blog, I feel that I am an important part of our classroom community.  
                      | 4. I have been stimulated to do additional readings or research on topics discussed on the blog.  
                      | 5. In comparison to my other classes, the amount of my interaction with other students in this class has increased due to the blog.  
                      | 6. In comparison to my other classes, the quality of interaction with other students in this class has increased due to the blog. |
| **Perceived learning**  | 1. The blog discussions help me to share my knowledge and experience with my peers.  
                      | 2. I believe that incorporating blogs with teaching can enhance my learning experience in general.  
                      | 3. Other students’ comments on my blog posts are important.         |
|                         | 4. Blog discussions help me understand other points of view.        |
|                         | 5. Blog discussions have made me think about nutrition concepts outside of this class.  
                      | 6. My point of view has been acknowledged by my peers and/or discussion leader in this course.  
                      | 7. Overall using the blog has helped me learn.                      |

### Procedure

The group project was designed to help students learn in a collaborative learning environment that incorporated blogging. We used the blog tool provided within Blackboard. The reasons to use blogs included ease of use and features of blogs that offer a collaborative learning space for each group. The groupwork was student-centered and project-based learning approach was undertaken in which the instructor served as a facilitator throughout the
process. The group project lasted one month. First, the instructor explained the group project process. Before students started groupwork, a short training session provided students with an overview of blog features. This was designed to decrease students’ fears of using blogs, especially for those who did not have any experiences blogging. Each group was asked to choose their own topic and post information relevant to the selected topic. In the last week of groupwork, students needed to review other groups’ blogs and provide comments. In addition, students were invited to participate in the survey after they completed the groupwork. The instructor provided guidance and directions to groups who encountered problems (e.g., collaboration procedure and topic selection) during the group project.

Data analysis

Data were analyzed using quantitative and qualitative methods (See Table 4). Quantitative approaches included descriptive analysis, correlation, regression analyses, and t-tests. Qualitative approaches included content analysis.

<table>
<thead>
<tr>
<th>Research questions</th>
<th>Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What are learners’ blogging self-efficacy, sense of community, perceived collaborative learning, and perceived learning?</td>
<td>Descriptive analysis</td>
</tr>
<tr>
<td>2. What is the relationship between learners’ blogging self-efficacy, sense of community, perceived collaborative learning, and perceived learning?</td>
<td>Correlation analysis</td>
</tr>
<tr>
<td>3. Do learners’ blogging self-efficacy, sense of community, and perceived collaborative learning predict perceived learning?</td>
<td>Regression analysis</td>
</tr>
<tr>
<td>4. Do learners who were aware of blogs before attending the class show better blogging self-efficacy and perceived learning than those who were not?</td>
<td>T-test analysis</td>
</tr>
<tr>
<td>5. What are learners’ perspectives of group learning experiences through the use of blogs?</td>
<td>Content analysis</td>
</tr>
</tbody>
</table>

Results

Descriptive information

Students possessed medium \((M = 3.83, SD = 0.88)\) blogging self-efficacy (See Table 5). The average score for sense of community was 3.94. Students perceived moderately high levels of collaborative learning with an average score of 4.21. Students perceived that blogging helped them learn, with a slightly high mean on perceived learning \((M = 4.31, SD = 0.50)\).

<table>
<thead>
<tr>
<th>Scales</th>
<th>Range</th>
<th>Midpoint</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogging self-efficacy</td>
<td>1-5</td>
<td>3</td>
<td>3.83</td>
<td>0.88</td>
</tr>
<tr>
<td>Sense of community</td>
<td>1-5</td>
<td>3</td>
<td>3.94</td>
<td>0.59</td>
</tr>
<tr>
<td>Perceived collaborative learning</td>
<td>1-5</td>
<td>3</td>
<td>4.21</td>
<td>0.43</td>
</tr>
<tr>
<td>Perceived learning</td>
<td>1-5</td>
<td>3</td>
<td>4.31</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Correlation and regression analyses

Self-efficacy \((r = .340, p < .01)\), sense of community \((r = .641, p < .01)\), and perceived collaborative learning \((r = .728, p < .01)\) were significantly correlated with learners’ perceived learning (See Table 6). Blogging self-efficacy was significantly correlated with sense of community \((r = .575, p < .01)\) and perceived collaborative learning \((r = .405, p < .01)\).

The multiple regression model (see Table 7) was significant, \(F(3, 56) = 30.34, p < .001\), explaining 61.9 % of the variance in perceived learning. Sense of community \((t(56) = 3.564, p < .05)\) and perceived collaborative learning \((t(56) = 5.509, p < .05)\) were significant predictors of perceived learning, with perceived collaborative learning the strongest predictor. Blogging self-efficacy did not significantly contribute to perceived learning, \(t(56) = -1.104, p > .05\).
Table 6. Correlations among variables

<table>
<thead>
<tr>
<th></th>
<th>Blogging self-efficacy</th>
<th>Sense of community</th>
<th>Perceived collaborative learning</th>
<th>Perceived learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogging self-efficacy</td>
<td>-</td>
<td>.575**</td>
<td>.405**</td>
<td>.340**</td>
</tr>
<tr>
<td>Sense of community</td>
<td>-</td>
<td>-</td>
<td>.557**</td>
<td>.641**</td>
</tr>
<tr>
<td>Perceived collaborative learning</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.728**</td>
</tr>
<tr>
<td>Perceived learning</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. **p < .01.

Table 7. Multiple regression model: Perceived learning explained by three predictor variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogging self-efficacy</td>
<td>-.063</td>
<td>.057</td>
<td>-.112</td>
<td>-1.104</td>
<td>.274</td>
</tr>
<tr>
<td>Sense of community</td>
<td>.338</td>
<td>.095</td>
<td>.399</td>
<td>3.564</td>
<td>.001**</td>
</tr>
<tr>
<td>Perceived collaborative learning</td>
<td>.631</td>
<td>.114</td>
<td>.551</td>
<td>5.509</td>
<td>.000***</td>
</tr>
</tbody>
</table>

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

**T-test analysis**

Students who were familiar with blogs before the class started had higher blogging self-efficacy than those who were not, t(58) = 2.096, p < .05, ES = 0.55. There were no significant differences in perceived learning between students who were familiar with blogs and those who were not (see Table 8).

Table 8. T-test analysis for blogging self-efficacy and perceived learning

<table>
<thead>
<tr>
<th></th>
<th>Knowing blogs</th>
<th>Not knowing blogs</th>
<th>t(58)</th>
<th>p</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogging self-efficacy</td>
<td>4.13</td>
<td>3.65</td>
<td>0.81</td>
<td>2.096</td>
<td>.040*</td>
</tr>
<tr>
<td>Perceived learning</td>
<td>4.41</td>
<td>4.24</td>
<td>0.52</td>
<td>1.241</td>
<td>.220</td>
</tr>
</tbody>
</table>

Note. *p < .05.

**Content analysis**

Student responses to their group learning experiences were visualized by using the word cloud application (see Figure 1). Descriptors (see Table 9) coded by the researchers were entered to obtain the visualized image. More frequently mentioned words or descriptors were shown in larger font size and brighter color.

Figure 1. Descriptions of learners’ perspectives of group learning experiences

Three themes emerged through coding: individual perceptions of groupwork, learning process, and collaboration process (see Table 9). Most students described group projects as positive, with very few students who disliked...
the group learning approach. Many students enjoyed groupwork and found the collaborative learning process enjoyable, effective, educational, and interesting. Some of them showed a high level of motivation to learn through collaboration.

The learning process covered the development of knowledge (e.g., content knowledge, thinking) and skills (i.e., organization skills, collaboration skills). It appears that through groupwork, more knowledge was learned than skills. Knowledge was developed through self-reflection that led to personal growth and changes of thinking. Real-life experiences helped students connect theory to practice and further develop their own understanding of important topics or issues.

Modes of collaboration reported by students included interaction, information sharing, and idea negotiation to achieve consensus. Idea exchange through discussion and communication helped eliminate disagreements among groupmates or overcome the challenges that took place during the collaborative groupwork. Feedback or resources shared from others in the same group made groupwork easier and also promoted idea formation through brainstorming.

Some students indicated that groupwork requires respect, patience, support, and encouragement from their groupmates. Feeling part of a team in which each group member’s ideas and contribution were valued was critical for group collaboration. A certain level of attention from groupmates was needed for some students, which helped them become more engaged in the collaboration process. Sense of responsibility and reliability was integral to keeping the whole interactive process smooth.

### Table 9. Major themes summarized from responses to the open question

<table>
<thead>
<tr>
<th>Themes</th>
<th>Categories</th>
<th>Descriptors</th>
<th>Selected examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual perceptions of</td>
<td>Positive (25)</td>
<td>• enjoying</td>
<td>“Our group activity has been an exciting adventure for me. I enjoyed working with each of my group members.”</td>
</tr>
<tr>
<td>groupwork (28)</td>
<td></td>
<td>• effective</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• exciting</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• educational</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• fun</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• great</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• highly-motivated</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• like</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• interesting</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• success</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative (3)</td>
<td>• dislike</td>
<td>“Its not effective to me to work in a group, I work better alone because I’m on my own time and I don’t have to wait for anybody. I work a lot so in my free time I like to work on my project(s) without trying to find out with anybody else have the same free time as I do.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• ineffective</td>
<td></td>
</tr>
<tr>
<td>Learning process (24)</td>
<td>Knowledge (21)</td>
<td>• knowledge acquisition</td>
<td>“I gained a [good] deal of knowledge from each of these ladies. I am so happy that I was able to “gleam” knowledge and information from them.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• personal growth</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• self-reflection</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• thinking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• real-life experiences</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• social-learning</td>
<td></td>
</tr>
<tr>
<td>Skills (3)</td>
<td></td>
<td>• organization skills</td>
<td>“I learned that sometimes you have to listen to other opinions in order to get the work done.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• collaboration-skills</td>
<td>“I really enjoyed my group because we bonded and learned to work together. I also learned that when working with a group, people will have many different ideas.”</td>
</tr>
<tr>
<td>Collaboration process</td>
<td>Group-oriented behaviors (64)</td>
<td>• agreements</td>
<td>“I learned that everyone will not agree on certain things. This group project has taught me that everyone has to have an opportunity to share their thoughts and ideas. While working in my group everyone came together</td>
</tr>
<tr>
<td>(82)</td>
<td></td>
<td>• assistance</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• collaborative</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• communication</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• decision-making</td>
<td></td>
</tr>
</tbody>
</table>
• disagreements
• discussions
• feedback
• idea-negotiation
• idea-exchange
• idea-sharing
• information-seeking
• information-sharing
• interaction
• participation
• resources
• various-opinions
• social-network
• challenges

Individual disposition toward team members (18)

• encouraging
• personalities
• reliable
• patient
• rewarding
• responsibility
• respect
• supportive
• team-feeling
• values
• contribution
• attention

“...and came to an understanding on the proper way to handle the assignment.”

“When participating in a group you should show acceptance to different personalities and be sensitive to people needs, feelings and positions.”

Note. The number within the parenthesis refers to the frequency of used descriptors.

Discussion

Blogging self-efficacy is positively related to perceived learning but not a significant predictor of perceived learning

Blogging self-efficacy was found to be positively correlated with perceived learning at a significant level. However, it did not significantly predict perceived learning. This may be because students had similar average blogging self-efficacy scores, which resulted in small variances among their perceived learning scores. Students may have underestimated the value of blogging and their motivation toward learning thus suffered (Wigfield & Eccles, 2000). Although a few previous studies have found blogging activities to have a positive impact on students’ grades (Xie, Ke, & Sharma, 2008; Xie, Ke, & Sharma, 2010), this study did not investigate the relationship between blogging self-efficacy and learning performance (e.g., grades); rather, it measured the association between students’ self-report of perceived learning (students’ self-rating of their perceived learning experiences) and blogging self-efficacy. Although suggested by the research of Top (2012), which involved pre-service teachers in ICT courses, such a relationship was not examined in any prior studies. Our study found that blogging self-efficacy is related to but not a critical predictor of perceived learning among minority students.

Taking a further look at the blogging self-efficacy scale, most students appeared to have good self-efficacy in using phones or digital cameras to take photos, shoot video, or transfer photos to a computer, but lower self-efficacy in terms of transferring video between different devices, and using a blog to read, comment or publish multimedia content. Students felt least confident using video- or image-processing software in blog activities. In this study, including images or videos was encouraged but not mandatory. Thus, it might be unlikely to find a significant influence of blogging self-efficacy on students’ perceived learning that assessed students learning experience using blogs to discuss, share knowledge, and learn.
Sense of community and perceived collaborative learning are significant indicators of perceived learning

Aligned with previous research, sense of community and perceived collaborative learning were critical to perceived learning (Halic et al., 2010; Kilic, & Gokdas, 2014; Top, 2012; Yang, 2009). They were not only significantly correlated with, but also significant predictors of, perceived learning. Previous studies did not involve African-American students, and this study provides evidence of such important relationships among minority populations. Using blogs helped students develop the feeling of belonging to a learning community, which is consistent with prior research (Muncy, 2014; Top, 2012; Top, Yulkelturk, & Inan, 2010; Wang, 2010; Yang & Chang, 2012). This makes sense in part because collaborative environments allow students to share information, exchange ideas, provide feedback, interact with others, and develop reflective thinking. Consistent with social constructivism, many educators have found blogs to be useful educational tools that provide social interactive environments through which students’ knowledge is created by conversing or interacting with others socially (Driscoll, 2005; Top, 2012; Top, Yulkelturk, & Inan, 2010; Wang, 2010; Yang & Chang, 2012). Such tools can support a variety of learning activities, especially for groupwork (Duarte, 2015).

Blogs are transformational technology tools that can potentially promote sense of community and perceived collaborative learning, which in turn, can enhance students’ perceived learning experiences in blog-supported learning environments (Top, 2012). According to the finding of this study, the higher the levels of sense of community and perceived collaboration, the better the perceived learning experiences for undergraduate minority students are. Students in this study who felt (a) more connected to their groupmates or (b) an important part of their group community were more likely to perceive better learning experiences, which aligns with previous findings (Halic et al., 2010; Top et al., 2010). Similarly, those who had more interaction or communication with groupmates were more likely to develop a better understanding of newly learned knowledge or skills (Papastergiou et al., 2011; So & Brush, 2008).

Prior knowledge of blogs may influence one’s blogging self-efficacy

Students who were familiar with blogs prior to the class had significantly higher blogging self-efficacy than those who reported no prior knowledge about blogs. This finding makes sense because the former students might have previously used certain blog features or creating their own blogs, which would possibly result in higher blogging self-efficacy than those who had not even known or utilized blogs. It is likely that blogging self-efficacy could be increased over time through relevant activities or other interventions that involve the use of blogs (Papastergiou et al., 2011). Previous researchers have not investigated the impact of prior blogging knowledge on blogging self-efficacy, and the significant finding of our study among minority students may be applicable to future blogging research.

Group collaborative learning contributes to the development of knowledge and skills

Most students perceived their groupwork experience as positive, which aligns with previous research. The group collaborative learning process with the use of blogs helped students gain both knowledge and skills through communication and idea sharing, which is aligned with the tenet of cultural-historical activity theory that knowledge is constructed through social interactions (Luria, 1976; Vygotsky, 1978). In terms of gaining knowledge, students perceived having learned the content, developed their ways of thinking through discussion, and gained real-life experiences.

Blogs in this study served as a mediating tool that provides students with collaborative learning environments and helps to shape their understanding of knowledge and acquisition of skills through meaningful communication or interaction over time, which reflects the sociocultural-historical perspective that highlights the importance of the interaction between tools, subjects, and objects (Hogan & Tudge, 1999; Vygotsky, 1978). Consistent with findings from previous studies, groupwork with blogging contributed to better understanding of course content (Halic et al., 2010; Li, Bado, Smith, & Moore, 2013) and the development of higher level thinking skills, such as theory-practical linkages, critical thinking and reflection (Deng & Yuen, 2011; Osman & Koh, 2013; Yang, 2009). Quality of peer comments plays an important role in critical or reflective thinking (Xie et al., 2008). Poor quality of peer feedback would have a negative effect on one’s reflective thinking. According to Osman and Koh (2013), collaborative blogging helped students to link theory to their professional experiences and their observations in the work place.
In terms of learned skills, students in this study learned how to manage tasks in an organized way and developed good collaboration skills. Such skills would be valuable for their current or future career. Prior research appeared to address more development of knowledge than that of skills in blogging groupwork. The two skills (i.e., collaborative and organizational skills) identified in this study were reasonable as group learning through blogging has shifted from instructor-centered to learner-centered and imposed greater responsibilities on students by requiring better management skills to achieve success (Garcia, Elbellagi, Brown, & Dungay, 2015).

**Individual dispositions are potential factors for collaboration processes**

The majority of students had positive reactions to collaborative groupwork, which is consistent with the results of previous research that indicates African-American students prefer to work collaboratively and learn better in such environments (Duncan & Barber-Freeman, 2008; Gallien & Peterson, 2004). However, none of these studies included the use of blogs, and no prior blogging studies have investigated African-American students’ learning experiences in blogging.

Individuals’ dispositions toward other members within the same group may influence the effectiveness of the collaboration process, as found in this study. This finding is valuable as no prior blogging research has focused on studying how individuals’ feelings, attitudes, or characteristics would possibly affect group collaboration in learning settings. Guadagno, Okdie, and Eno (2008) researched students from a university and found that those who had high levels of openness as well as neuroticism were most likely to be bloggers. People who maintained personal blogs were most likely to use other online sources or platforms for social interaction. Similarly, extroverted people are more likely to have more interactions with others than introverted people in online learning (Kuo, Walker, Belland, Schroder & Kuo, 2014). The important factors found in this study, such as patience, reliability, responsibility, respect, support, etc., might be associated with the characteristics (e.g., extroversion, high openness, high neuroticism) identified among most bloggers.

**Conclusion**

Overall, the minority students in this study were positive about the use of blogs in groupwork, and most of them preferred working in a learning community. Sense of community and collaborative learning significantly predicted perceived learning, with collaborative learning being the strongest predictor. This provides evidence of the usefulness of using blogs in enhancing minority students’ collaborative learning experiences and sense of belonging in a group. Blogging self-efficacy, a predictor suggested to be included but not investigated by previous researchers, was not found to significantly predict perceived learning, although previous knowledge of blogging was positively related to blogging self-efficacy at a significant level. Content analysis showed that most students liked groupwork and that their knowledge and skills (i.e., management skills, organization skills, and collaboration skills) were developed throughout the group learning process. Group-oriented behaviors (e.g., communication, information seeking) were largely identified through collaboration processes. Individual positions, including peer support from and emotional bonds among groupmates, were particularly important during collaboration.

This study not only confirms the preference of collaborative group work among minority students as well as the importance of collaborative learning and sense of community on perceived learning in blog-enhanced settings, but also adds to the limited research about the influence of blogging self-efficacy on perceived learning, as well as the effectiveness of web 2.0 tools to enhance knowledge and skill acquisition, collaboration, and affective learning outcomes in the classroom among non-traditional minority students in continuing education. Implications suggest that course instructors should (a) identify students who are not familiar with blogs and provide them with additional training before groupwork; (b) design small icebreaker activities to help members in the same group know each other and to develop sense of learning community; (c) apply proper pedagogies or strategies, such as case-based, project- or problem-based learning, to enhance student collaboration in an effective way; and (d) include learners’ real-life experiences in groupwork to facilitate collaborative learning. In the future, researchers are encouraged to (a) replicate this study among other students from minority backgrounds; (b) examine the relationship between blogging self-efficacy and perceived learning with different student populations; (c) further study the association of levels of collaboration and sense of community with learning performance, and (d) investigate how individual attributes, characteristics, or emotional variables have an influence on the collaborative process.
In terms of limitations of this study, the findings may not be generalized to other ethnic populations or traditional on-campus students. This study involved non-traditional adult students who were more mature than traditional on-campus students. The sample was drawn from two classes, which may not well present all African American students. In addition, we adopted the original instruments developed by previous researchers and did not further examine the validity of these scales with our sample due to the small sample size (Arrindell & van der Ende, 1985; Myers, Ahn, & Jin, 2011). Future studies are encouraged to further validate these scales with different populations or contexts, and examine their reliability to ensure the consistency of items.

References


Using Visualization to Motivate Student Participation in Collaborative Online Learning Environments

Sung-Hee Jin

Frontier College, Inha University, Nam-gu, Incheon, South Korea // shjin@inha.ac.kr

(Submitted December 29, 2015; Revised May 30, 2016; Accepted May 31, 2016)

ABSTRACT

Online participation in collaborative online learning environments is instrumental in motivating students to learn and promoting their learning satisfaction, but there has been little research on the technical supports that motivate learners. The purpose of this study was to develop a visualization tool to motivate learners to participate actively in collaborative online learning communities and examine its effects on online participation, perceived learning, perceived satisfaction, team project performance, and usability. Two types of visualization tool were developed: the A-type for representing group participation and interaction between groups, and the B-type for representing group participation and individual participation. Undergraduate students ($N = 118$) were assigned randomly to one of three groups (A and B: experimental group, C: control group). The results showed that two types of visualization tool have significant effects on the perceived satisfaction, perceived learning, and team project performance. Regarding online participation, the B-type of visualization tool was found to be effective. The usability of the A and B groups was similar and many learners in both groups reported that the visualization tool motivated them to participate in online learning. Theoretical, empirical and practical discussions are provided based on the results.

Keywords
Visualization of participation, Online participation, Visual scaffolding, Visual feedback, Participation motivation

Introduction

Research has suggested that online participation, including social interactions in collaborative online learning environments, is instrumental in motivating the students to learn and promote their learning satisfaction (Cobb, 2009; Hrastinski, 2009). In face-to-face learning environments, learners can naturally know who is an active participant in learning and which group is more collaborating, and subsequently be motivated to participate more by them. On the other hand, engaging students to participate in online communities is not easy because many learners can be indolent and contribute very little (Barria, Scheihing, & Parra, 2014). Learners in collaborative online learning environments need participation information of individuals and groups (Gross, Stary, & Totter, 2005). Providing this participation information to learners can support online participation and learning outcomes (Nilsson & Svensson, 2012). A few previous studies have reported that visualization tools for individual or group participation have potential effects on learning participation, collaborative activities, or group performance in online communities (Janssen, Erkens, & Kirschner, 2011; Hsiao et al., 2012; Sun & Vassileva, 2006). These studies suggested the applicability of a visualization tool on online participation; however, they had a limitation in that visualization tools were not properly designed based on theoretical backgrounds, and participation data for visualizing was used in a restricted manner. The aim of this study was to develop a visualization tool for participation information to motivate learners to participate actively in collaborative online learning environments. The effects of a visualization tool were also investigated in terms of online participation, perceived learning, perceived satisfaction, team project performance, and usability.

Online participation and visualization

In order to achieve the purpose of this study, online participation was defined and the limitation of the previous studies related to visualization of online participation was identified. A visualization of online participation can help learners aware their participation level, so the types of awareness information were reviewed and the useful guidelines for visualizing information were summarized.

Online participation

Hrastinski (2008, p. 1761) defined participation as “online participation is a process of learning by taking part and maintaining relations with others.” Online participation involves many types of engaging activities, such as...
talking, writing, reading, watching, or thinking. These online participation events can be collected and analyzed in collaborative online learning environments. Many studies have found empirically that online participation has positive effects on the learning satisfaction and retention as well as the learning outcomes (Alavi & Dufner, 2005; Morris, Finnegan, & Wu, 2005). After a survey targeting 1406 online learners, the State University of New York reported that the most important factors influencing the online learning effectiveness were (1) interaction with the teachers, (2) level of participation, and (3) interaction with classmates (Frederiksen et al., 2000). Morris et al. (2005) also proposed that online participation has a positive effect on the learning outcomes. Online participation in collaborative learning environments can be defined operationally as a process of learning by accessing a range of learning environments, maintaining relationships with peers and instructors, performing individual or group work, and sharing ideas or learning materials.

Visualization of online participation

A few studies have investigated the effects of visualizing individual or group online participation on learner participation or learning outcomes. Sun and Vassileva (2006) created motivational community visualizations with static and dynamic versions in peer-to-peer file-sharing networks and examined their effects on online participation and awareness of the online learning community. They did not find any evidence that the static visualization encouraged participation, however dynamic visualization effectively increased the learner’s participation and awareness of the online learning community. Dynamic visualization could be viewed according to topic, number of original postings, number of shared files, login-frequency, and total participation. Janssen, Erkens, Kanselaar, and Jaspers (2006) examined the effects of learner participation visualization on the individual participation, awareness, collaborative activities, and group performance scores in an online chat room. The Participation-tool (PT) visualized the average length and the number of messages with spheres and lines, respectively. As a result, the PT increased the average length of the messages but had no effect on the awareness and group performance scores. This suggests that the size of a sphere, which indicates the average length of the messages sent by a student, makes the learner write longer messages. Janssen et al. (2011) examined again the effects of the PT on online collaboration and reported that the duration of PT use has a significantly effect on the group members’ participation and equality of participation within the group, but they have no effect on the group performance. Stankiewicz (2015) also developed a visualization for students’ writing and viewing behaviors in collaborative environments and examined its effects on the quantity of collaborative behaviors. He did not find sufficient evidence that the visualization led to more collaborative behaviors because the visualization might be unclear and ambiguous according to the student interview. Although the results of previous research related to visualization tools for online participation have been inconsistent, the potential effects of visualizing the online learning participation can be explained in terms of the awareness information.

Awareness information

Awareness is defined as “understanding of the activities of others, which provides a context for your own activity” (Dourish & Bellotti, 1992). Awareness information in collaborative online environments plays a role as making one’s activity visible to others (Dourish, 1997). Nilsson and Svensson (2012) suggested that awareness information on participation can support participation and learning in online learning communities. Two types of awareness, group awareness and objective self-awareness, which were suggested by Gross, Stary and Totter (2005), can be considered when designing a visualization tool for online participation. Group awareness gives an overview of group members’ roles, activities, movements and status in the collaborative process (Greenberg, Gutwin, & Cockburn, 1996). The following three types of group awareness can be considered as being crucial for effective collaborative learning: behavioral awareness provides information on the learners’ activities in collaborative learning environments (e.g., Janssen et al., 2011); cognitive awareness reveals the knowledge level of the group members (e.g., Dehler, Bodemer, Buder, & Hesse, 2011); and social awareness reveals the functioning of the group, as perceived by the collaborators (e.g., Phielix, Prins, Kirschner, Erkens, & Jaspers, 2011). Objective self-awareness means the process of taking oneself as the focus of one’s own attention, or becoming aware of oneself (Mullen & Goethals, 1987). This awareness addresses the individual performance, i.e., if one is sufficiently aware of oneself, one may come to recognize a discrepancy between one’s present behavior and the standard of behavior. The individual attempts to reduce the discrepancy depending on his/her outcome expectancy (Mullen, Migdal & Rozell, 2003).
Information visualization

The general guidelines for visualizing an online community have been suggested in the field of information visualization. Erickson and Kellogg (2003) emphasized that the information shown in the visualization does not need to be detailed and precise. All learners should see the same thing so that they can feel responsible for their actions because they know that others will see the same things and be aware of what they do (Ericson, 2003). The choice of metaphor is also very important because the learners should not put their efforts in interpreting a visualization tool. A metaphor should be intuitive to use and not require cognitive overload for interpretation (Sun & Vassileva, 2006). Appropriate use of the location and color contraction of the visual components will also attract attention (Lamme, 2003).

Design directions for a visualization tool

The matrix shown in Table 1 organizes the design implications and directions for a visualization tool on online participation according to the theoretical foundations (i.e., concept of online participation, awareness information and information visualization).

<table>
<thead>
<tr>
<th>Theoretical foundations</th>
<th>Design implications</th>
<th>Design directions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online participation</td>
<td>Online participation includes - accessing to the learning environments - maintaining relations with peers and instructors - performing individual or group work - sharing ideas or learning materials</td>
<td>(1) Collect, analyze and visualize data to assess the level of online participation in a collaborative learning management system. - login frequency, number of original postings/responses &amp; comments/read messages/votes, message lengths,</td>
</tr>
<tr>
<td>Awareness information</td>
<td>(Group awareness) Being informed about the specific aspects of group members strongly affects their collaboration (Objective self-awareness) Recognizing a discrepancy between one’s present behavior and the standard behavior make one try to reduce the discrepancy</td>
<td>(2) Visualize the level of online participation of group members</td>
</tr>
<tr>
<td>Information visualization</td>
<td>The information shown in a visualization tool does not need to be precise</td>
<td>(3) Visualize the relative status on online participation</td>
</tr>
<tr>
<td></td>
<td>All should see the same thing</td>
<td>(4) Design to provide a general concept rather than detail information on online participation</td>
</tr>
<tr>
<td></td>
<td>A metaphor should not require cognitive overload for interpretation</td>
<td>(5) Provide the same visualization tool to all the learners</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6) Use an intuitive and understandable visual metaphor</td>
</tr>
</tbody>
</table>

This study attempted to visualize the online participation with critical measures considering the concept of online participation in collaborative online learning communities, which was suggested in this study, to overcome the limitations of previous research. The quantitative data for visualizing the online learner participation was only designed to make a visualization tool easily reusable in similar situations. The qualitative data was mostly analyzed manually because there has been little research on the implementation of analysis tools of qualitative data in a learning management system. The following units of online participation analysis were initially selected: login frequency, number of original postings, number of responses and comments, message lengths, read messages, and time spent. Expert reviews were conducted to validate the six types of online participation behaviors. Five experts participated in two expert reviews; they had a PhD. in educational technology and had more than five years’ experience in studying and teaching online courses. The purpose of this study was introduced to the experts by telephone and an expert review instrument was provided and collected by e-mail. The instrument for the expert review was designed to evaluate the appropriateness of the selected units for online learner participation using a four-point Likert-type scale. The instrument was composed of the following 7 items: six for the appropriateness of the selected analysis units and one asking for any comments on the units. The validity of the expert responses was analyzed using the content validity index (CVI) and the inter-rater agreement (IRA) adopted in Jin (2013). According to the results of the first expert review, all
CVI and IRA values for the five types of analysis units exceeded 0.8 except for the “time spent” unit. Two experts reported that the time spent unit could be biased in many cases. Three experts suggested adding the unit of the voting number to the participation and interaction value.

Therefore, a second expert review was conducted to validate the revised units of the analysis and explore their weighted value using the same method as the first expert review. The instrument was composed of 7 items: six for the appropriateness of the revised analysis units and one for asking to write the weighted values in each unit. All CVI and IRA values for the six types of analysis units exceeded 0.8, suggesting that all units were valid and highly reliable. The weighted value of each unit was first yielded according to the mean value of the experts’ opinions and were determined as follows by another experts review: 1 for the login frequency, 3 for the number of original postings, 2 for the number of responses and comments, 2 for the message lengths, 1 for the read messages, and 1 for number of votes.

Referring to the design implications based on the awareness information, a visualization tool should represent the group participation as well as the individual participation. Moreover, it should represent the relative status of online participation. Therefore it is intended for the learner to make efforts to improve the individual or group participation.

Representing the overall information on online participation is sufficient in designing a visualization tool. Learners just need to be provided with a general concept for their participation. Erickson and Kellogg (2000) related the concept of social translucence; they addressed ‘accountability’ in collaborative online communities. Accountability is related to the fact that they also know about you, as you know about others in a socially translucent system. Therefore, all the learners should be provided with the same visualization tool. Lastly, a visual metaphor following these design directions is important. A method to represent general social networks can be applicable and useful.

**A visualization tool on online participation**

A visualization tool represents the individual participation, group participation and the interaction among individuals or groups considering the characteristics of collaborative online learning. Representing the individual participation and interaction was intended to provide information on the learner’s objective self-awareness. Representing the group participation and interaction was intended to inform learner group awareness. To determine which information is critical for online learning, two types of visualization tools were developed to reflect the characteristics of the awareness information represented. The A-type represents the group participation and the interaction between groups that identify the group awareness information, and the B-type represents the group participation and individual participation that identifies the group and objective self-awareness information. The online participation data is displayed with circles, squares and lines referring to the representation method of various social networks.

![Figure 1. Screenshot from the A-type of visualization tool: (a) group interaction levels; (b) group participation levels](image)

Figure 1 presents a screenshot of the A-type of visualization tool representing the group participation and group interaction. The color and size of each square represents the level of group participation. The group participation.
The group interaction value was calculated from the number of follow-up posts, number of reading posts, and number of votes on the posts among the members of the two groups. The interaction value also reflects the weights of the target analysis units. Ten different color codes for the group participation level and five different color codes for the group interaction level were used.

Figure 2 presents a screenshot of the B-type of visualization tool representing the group and individual participation. The color and size of the circles represent the level of individual participation, so the learner’s name is presented in each circle. The color and size of the squares in the middle of each group represent the level of group participation. The individual participation value was calculated using the data from six analyses and their weighted values. The group participation value was produced using the same method applied in the A-type of visualization tool. The interaction value was not represented in this visualization tool. Ten different color codes for the individual and group participation level were used.

An expert validation test was conducted to determine if the individual participation, group participation, and group interaction were represented appropriately in the developed visualization tool. The same experts who took part in the former expert review for the analysis units selection participated. Using a four-point Likert-type scale (strongly disagree/strongly agree), the five experts were asked about the implementation appropriateness on the individual participation, group participation, and group interaction. As a result, all experts responded with at least three points.

Research method

The visualization tools were validated by controlled testing based on a development research methodology. An experimental study was carried out to determine the effects of the visualization tool on online learning. The specific research questions were as follows:

- To what extent do the visualization tools influence the learner’s online participation?
- To what extent do the visualization tools influence the learner’s perceived learning, perceived satisfaction, and team project performance?
- Are there any differences in usability between the groups that received an A-type or B-type of visualization tool?
- What do the learners think when they look at the visualization tools?

Participants and design

The participants were 118 undergraduate students (91 males, 27 females) enrolled in a “Creative Thinking” course at a university in South Korea. All students came from an engineering college and were assigned...
randomly to one of three groups. Forty three students (30 males and 13 females) were included in the first experimental group, which received the A-type of visualization tool representing the group participation and group interaction. Forty students (32 males, 8 females) were assigned to the second experimental group, which received the B-type of visualization representing the group participation and individual participation. Thirty five students (29 males, 6 females) were included in the control group, which used the same online learning community as the experimental group learners but they did not receive any type of visualization tool. The experiment was carried out as part of the student’s regular coursework. Six units of online participation analysis were not explained to groups A or B. Thus, the learners in both groups sometimes asked the teachers how to yield the participation or interaction levels, and the teachers always answered that the visualization tool represented synthetically their online participation.

**e-Learning activities in the “Creative Thinking” course**

The course, “Creative Thinking,” was designed to enhance the divergent thinking ability of engineering students. This course provides various team-based activities for the learner to explore and develop his/her creative thinking. This course is normally operated by blended learning. For in-class activities, a range of creative thinking methods are explained and the learners apply the learned methods to solve problems. For online activities, the teacher gives the learners a real-world problem to solve with the acquired creative thinking methods each week. The grades are determined by the in-class and online activities. A new online learning community was created for this study. The students did not provided a visualization tool for the initial eight weeks to allow them to become accustomed to using the new online learning community. The results of examining the homogeneity among groups showed no significant differences in the levels of online participation (Group A: M = 30.17, SD = 16.27, Group B: M = 33.56, SD = 11.48, Group C: M = 33.71, SD = 18.14) and interaction (Group A: M = 51.39, SD = 26.64, Group B: M = 45.72, SD = 32.17, Group C: M = 66.91, SD = 47.64). The experiment study was then performed over a seven week period. The students had to perform the team-based project activities and submit the outputs of each process every week through a design thinking process in the online community. Any teaching strategy to support collaborative activities did not applied, because the strategies can be biased to examine the effects of the visualization tools. Two teachers taught three classes with the same learning materials and evaluation plans. One taught the first experimental group and the other taught the second experimental group and control group.

**Measures**

The measures for this study consisted of online participation, perceived learning, perceived satisfaction, usability test, students’ opinions and team project score. The online participation was examined with six analysis units (login frequency, number of original postings, number of responses and comments, message lengths, read messages, and number of votes) and an influence test of the visualization tools. The data from six analysis units was collected from web-log data. The influence test (a = 0.81) of the visualization tools was intended to measure how much the visualization tools influenced the learner’s online participation, such as login to the system, writing original posts, writing follow-up posts, facilitating group members’ online participation, and accessing the visualization tool. The perceived learning test and perceived satisfaction test items were developed by referring to Eom, Wen, and Ashill (2006) and Kang and Park (2010) to measure the degree of students’ perceived learning and satisfaction through their online learning activities. The perceived learning test consisted of four items (a = 0.81), and the sample item was “I feel that I have learned a lot from this online learning activities.” The perceived satisfaction items (a = 0.85) were three and the sample item included “I would recommend this course to other students.” The usability test was developed by referring to the items suggested by Nokelainen (2006) and was assessed using the following three measuring scales: learnability (2 items; a = 0.78), effectiveness (2 items; a = 0.76), and satisfaction (3 items; a = 0.84). An open-ended question was provided to the students in the two experimental groups to identify the students’ opinions on the visualization tool. The validity of all test items was examined by two of the five experts who participated in the previous two experts’ reviews. Lastly, the team project scores were then yielded using an evaluation rubric for assessing the engineering design activities. Two teachers evaluated all team project reports of the three groups and the team project score in each team was the mean value between teachers. All tests was performed in a classroom under a teaching assistant’s guidance after completing the course.
Data analysis

Multivariate analysis of the variance was conducted to determine the effects of two types of visualization tool on the learner’s online participation, perceived learning, perceived satisfaction, and usability. The data was examined for normality, presence of outliers, and homogeneity of the variances before performing MANOVA. The partial eta-squared ($\eta^2$) and Cohen’s $d$ were calculated to determine the effect size if a significant multivariate result was found. The partial eta-squared values of approximately 0.01, 0.06, and 0.14 indicated small, medium, and large effects, respectively (Richardson, 2011). Cohen’s $d$ can be interpreted as follows: small < 0.2, medium = 0.5, and large > 0.8 (Cohen, 1988). The mean values of the team project performance were compared among the groups due to the insufficient number of teams. The learners’ responses of an open-ended question were analyzed using a content analysis method (Krippendorff, 2004). Each response was reviewed on a line-by-line basis, and the units of meaning were identified. The units consisted of words, phrases and/or sentences that contained meaningful information about the learner’s thoughts on the visualization tool. The units were coded and grouped into three categories; affective, cognitive and aesthetic opinions. The results of content analysis were cross-checked for validity by two PhD qualified educational technologists. A cross validation check revealed a proportional agreement of 0.92 between the analysts. Disagreements were resolved through discussion. Frequency analysis was conducted to compare the learners’ responses based on the units analyzed among the groups.

Results

Does the visualization tool encourage learner online participation?

Two types of data analysis were conducted to determine if the visualization tools encourage learner online participation. First, six sets of analysis data from the web-log database were extracted and the questionnaire data on the influence test of the visualization tool was then assessed. Table 2 lists the means and standard deviations of the three groups regarding the online participation (individual participations, peer interactions, and six type of participations data). The mean values of the dependent variables related to online participation of the two groups that received the A- or B-type of visualization tool were higher than those of group C, which did not receive any type of visualization tool. MANOVA revealed significant effects for the B-type of visualization tool on individual participation ($F(2, 115) = 13.877, p = .000, \eta^2 = .209$) and peer interactions ($F(2, 115) = 18.653, p = .000, \eta^2 = .262$). On the other hand, no significant effect was observed for the A-type of visualization tool, which represented the group participation and interaction. In particular, the B-type of visualization tool had positive effects on the login-frequency ($F(2, 115) = 4.836, p = .010, \eta^2 = .084$), number of follow-up posts ($F(2, 115) = 22.073, p = .000, \eta^2 = .296$), and message length ($F(2, 115) = 14.345, p = .000, \eta^2 = .215$) compared to those of the other two groups. Regarding the number of reading messages, a significant effect was observed for the two types of visualization tools ($F(2, 115) = 9.096, p = .000, \eta^2 = .148$).

| Table 2. Comparison of the online participation among the three groups |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
|                             | Group A                     | Group B                     | Group C                     |
|                             | A-type of visualization tool | B-type of visualization tool | Control group               |
|                             | (n = 43)                    | (n = 40)                    | (n = 35)                    |
|                             | M    | SD | M    | SD | M    | SD |
| Participations              | 40.13 | 20.87 | 56.16 | 19.85 | 33.86 | 12.81 |
| Interactions                | 87.65 | 46.41 | 228.76 | 184.04 | 77.74 | 64.25 |
| login frequency             | 44.07 | 24.46 | 59.41 | 35.17 | 40.67 | 18.14 |
| num.original posts          | 8.44  | 4.51 | 8.61 | 2.45 | 8.29 | 3.12 |
| num.follow-up posts         | 5.31  | 4.27 | 30.59 | 32.26 | 3.26 | 4.56 |
| message lengths             | 1324.68 | 1057.80 | 2308.00 | 1164.36 | 1178.58 | 540.03 |

**F value**

- $13.877$ **(P-value 0.000)**
- $18.653$ **(P-value 0.000)**
- $4.836$ **(P-value 0.010)**
- $4.076$ **(P-value 0.027)**
- $22.073$ **(P-value 0.000)**
- $14.345$ **(P-value 0.000)**

**Post hoc test**

- $A > B(0.76)$
- $A > C(1.74)$
- $A > B(3.04)$
- $A > C(2.35)$
- $A > B(0.62)$
- $A > C(1.03)$
- $A > B(5.92)$
- $A > C(5.99)$
- $A > B(0.92)$
- $A > C(2.09)$
MANOVA also revealed significant effects of the B-type of visualization tool on the influence ratings and each of the four influence scales: login to the system, $F(2, 115) = 19.509, p = .000, \eta^2 = .216$; writing original posts, $F(2, 115) = 15.344, p = .000, \eta^2 = .178$; writing follow-up posts, $F(2, 115) = 19.443, p = .000, \eta^2 = .215$; and access to the visualization tool, $F(2, 115) = 11.900, p = .001, \eta^2 = .144$. These results mean that the learners in group B login to the system more, and write more original and follow-up posts after receiving the visualization tool than those in group A.

**Does the visualization tool enhance the learner’s perceived learning, perceived satisfaction and team project performance?**

The effects of the visualization tool were significant on the learners’ perceived learning ($F(2, 115) = 6.206, p = .003, \eta^2 = .109$), and perceived satisfaction ($F(2, 115) = 3.998, p = .021, \eta^2 = .073$). Regarding the team project performance, MANOVA could not be conducted because of the insufficient number of teams but the results showed that the mean values of groups A and B, which received the visualization tools, were higher than those of group C, which did not receive a visualization tool.

**Are there any differences in usability between groups which received A-type or B-type of visualization tool?**

As shown in Table 3, the mean ratings of group B were higher on the usability tests than those of group A. MANOVA showed that the usability test and its sub-scales had no significant effect except for the effectiveness between the two types of visualization tool. The effect of the B-type of visualization tool reached statistical significance for the effectiveness scales ($F(2, 115) = 6.662, p = .012, \eta^2 = .092$).

**Table 3. Mean score and standard deviation of the three groups on the dependent variables**

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Group A</th>
<th>Group B</th>
<th>Group C</th>
<th>$F$ value</th>
<th>$P$ value</th>
<th>Post hoc test (effect size)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A-type of visualization tool ($n = 43$)</td>
<td>B-type of visualization tool ($n = 40$)</td>
<td>Control group ($n = 35$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Perceived learning</td>
<td>4.11</td>
<td>.64</td>
<td>4.00</td>
<td>.61</td>
<td>3.54</td>
<td>.80</td>
</tr>
<tr>
<td>Perceived satisfaction</td>
<td>4.04</td>
<td>.77</td>
<td>4.03</td>
<td>.79</td>
<td>3.54</td>
<td>.87</td>
</tr>
<tr>
<td>Team project performance</td>
<td>14.45</td>
<td>-</td>
<td>15.4</td>
<td>-</td>
<td>12.3</td>
<td>-</td>
</tr>
<tr>
<td>(teams)</td>
<td>(11)</td>
<td></td>
<td>(10)</td>
<td></td>
<td>(9)</td>
<td></td>
</tr>
</tbody>
</table>
What do learners think when they look at the visualization tool?

The learners’ thoughts that come to mind when they look at the visualization tool can be categorized into affective, cognitive and aesthetic opinions (See Table 4). Most learners in Groups A and B thought that the visualization tool motivated them to participate in online learning through a competitive comparison. The number of learners in Group B who reported the visualization tool’s motivation effect was higher than that in Group A (Group A: 13 vs. Group B: 27). This concurs with the result of research problem one in this study (See Section Does the visualization tool encourage learner online participation?). In addition, the B-type of visualization tool motivates the learners to participate in online learning better than the A-type. Regarding the cognitive aspect, the learners gave their opinions that the visualization tool helped them understand their online participation degree intuitively and made them self-reflect their online learning. The number of learners in Group A, who reported they felt unpleasant when they looked at the visualization, was higher than those in Group B. On the other hand, group B had more learners with negative feelings than Group A. Finally, a few learners in both groups reported that the visualization tool provided was fresh and interesting.

Table 4. Learners’ opinions on the visualization tool

<table>
<thead>
<tr>
<th>Opinions</th>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Affective</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivate to participate in online learning due to comparison</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>Positive feeling: great and pride</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Unpleasant feeling: envy, regrettable, sad, sorry</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>Negative feeling: bad, inferior, and unease</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td>23</td>
<td>32</td>
</tr>
<tr>
<td><strong>Cognitive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy of understanding the degree of online learning participation at a glance</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Self-reflection on his/her own online learning</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Effective method for facilitating online participation</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Effective for upper and middle groups but negative for lower groups</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td><strong>Aesthetic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresh and interesting</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Discussion

Theoretical implications

The instructional effects of the visualization tool on online participation can be explained theoretically in terms of the awareness information, visual feedback, and social comparison. Schmidt (2002) suggested two types of
important activities, displaying and monitoring, for being aware group or oneself in collaborative online learning environments. The visualization tool may help the learners objectively monitor their status of online participation by displaying their online participation. Although the learners perform team-based activities, they are motivated by the level of individual participation more than the level of group participation. Namely, this study found that objective self-awareness information has a more positive effect on individual participation than group awareness information. These results are also supported by Janssen’s et al. (2006) study, who suggested that the visualization of participation can affect participation through feedback processes. The learners want to receive comments or feedback on what they have done. The visualization tool makes the learners aware and monitors their online participation by providing visual feedback as to what degree they participate in online learning. The visual feedback helps the learners self-evaluate their online learning and makes them compare their participation levels with others. As a result, visual feedback and social comparison motivate the learners to participate more actively. Festinger’s (1954) theory of social comparison states that an upward comparison process refers to viewing others performing marginally better than themselves. This might encourage the learners to set higher personal standards that can motivate them to improve themselves. The visualization tool may help learners make an upward comparison, so learners may be motivated to increase their participation.

The effects of the visualization tool on online participation were similar to the results reported by Sun and Vassileva (2006), and the effects on the learning outcomes were consistent with Michinov and Primois (2005). Janssen et al. (2006) visualized the learner participation and reported that the participation tool increased the message length but did not affect the other dependent variables. The results reported by Janssen et al. (2006) may have been affected by the size of the sphere in the visualization tool determined from the average length of the learner’s messages. Therefore, the visualization tool in this study was designed with various learner’s participation data and the learners were not given a specific explanation. These uncertainties on the visualization tool may have a positive effect on their online participation.

**Design implications**

The differences between the current implementation and the previous participation visualizations in collaborative learning environments can be discussed in three aspects: using various participation data according to the concept of online participation, designing the visualization tools based on theoretical backgrounds, and using visual metaphor for learners to understand easily. This study provides an effective design strategy that makes the learners self-reflect and promotes online participation. In addition, it makes the teachers understand the learners’ online participation and become aware of who requires the teachers’ support. More importantly, the visualization tool suggested in this study is easy to understand. Table 1 lists the specific design guidelines for visualizing the student’s participation in collaborative online learning environments.

**Limitation and future directions**

This study examined the effects of a visualization tool on online participation, perceived learning, perceived satisfaction, and team project performance. The levels of online participation were determined from quantitative online participation data, such as the login frequency, number of original postings, number of responses and comments, message lengths, messages read, and number of votes. As suggested by Richards (2011), in that the learner’s engagement cannot be explained fully by a small amount of quantitative data, future research will be needed to represent the learner’s qualitative online participation and examine its effects on online learning. Moreover, the effects of the visualization tool may differ according to the learner’s characteristics. In particular, future research will be needed to determine if there are differences in the effects of the visualization tool according to the learners’ social comparison motive. To understand the effects of the visualization tool in more detail, the learner’s thoughts or feelings on the visualization tool need to be analyzed using a qualitative research methodology, and the learner’s attention or eye movement needs to be analyzed using an eye tracking method. In addition, the visualization tool initially attempted to represent the individual interactions and individual contributions to the group participation, however it could not be represented due to technical limitations. Therefore, future research will be needed to develop a visualization tool representing the individual interactions and individual contributions to the group participation, and examine its effects.
Conclusions

These results show that the visualization tool on online participation can enhance their online participation. In particular, the visualization of individual participation has greater effects on the online participation than the visualization of group participation despite the collaborative learning community. These findings will help guide the design of collaborative online learning communities to motivate learners to participate by making them reflect on their online learning participation themselves without the need for coaching or guiding of teachers or tutors. Moreover, the visualization tool can be adapted easily to collaborative online learning environments because of the ease of development.

Acknowledgements

This research was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2013R1A1A2063928 & NRF-2015S1A5A2A03048269).

References


Test Anxiety Analysis of Chinese College Students in Computer-based Spoken English Test

Yang Yanxia
The College of Foreign Studies, Anhui Xinhua University, Hefei, Anhui, China // yanxiyang3790@163.com

(Submitted March 9, 2016; Revised June 26, 2016; Accepted July 17, 2016)

ABSTRACT
Test anxiety was a commonly known or assumed factor that could greatly influence performance of test takers. With the employment of designed questionnaires and computer-based spoken English test, this paper explored test anxiety manifestation of Chinese college students from both macro and micro aspects, and found out that the major anxiety in computer-based spoken English test was spoken English test anxiety, which consisted of test anxiety and communication apprehension. Regard to proximal test anxiety, the causes listed in proper order as low spoken English abilities, lack of speaking techniques, anxiety from the evaluative process and inadaptability with computer-based spoken English test format. As to distal anxiety causes, attitude toward learning spoken English and self-evaluation of speaking abilities were significantly negatively correlated with test anxiety. Besides, as test anxiety significantly associated often with test performance, a look at pedagogical implications has been discussed in this paper.

Keywords
Computer-based spoken English test, Test anxiety, Test anxiety causes

Introduction

With development of technologies and large-scale testing requirement, the need for applying computers in language testing was inevitable in academic contexts (Hosseini, Abidin, & Baghdarma, 2014). In 1999, Chinese Examinations Board of CET-4 (College English Test-Band 4) and CET-6 (College English Test-Band 6) officially launched Spoken English Test (SET), a typical face-to-face test approach, to examine college students’ speaking abilities. However, as the face-to-face spoken English test took much time, effort and expense, which usually resulted in a limited number of test takers, in 2005 the National College English Testing Committee proposed to adopt Computer-based Spoken English Test (CBSET) to measure and assess college students’ speaking abilities and communicative skills. Computer-based test featured high-tech and interactive questions that incorporated diverse test stimuli and objectiveness, like video and graphics, and superiorly recorded voice quality (Davis, 2012). However, when a person was confronted with a novel testing situation, anxious emotion would probably arise and influence his or her test performance (Spielberger & Vagg, 1995; Wenemark, Persson, Brage, Svensson & Kristenson, 2011). With the prevalence of developing and using computer-based tests in educational assessment, issue of test anxiety resulted from the new testing approach of employing computers has raised concerns of researchers.

Literature review

Test achievement was very important and decisive in a person’s academic development. In order to achieve high test scores, students are naturally under great pressure. Similarly, results obtained to assessment were usually taken into account when decisions to be made with regard to individuals in different environments, hence assessments have become stimuli that caused anxiety reactions (Zeidner, 1998). When students took a test, their performance would be expected to be influenced by the perceived consequence of the test (Wolf, Smith,& Birnbaum, 1995). Thus, in the search for approaches to reduce test anxiety, researches have to first focus on identifying antecedents to test anxiety (Jain & Dowson, 2009; Zeidner & Matthew, 2005).

Test anxiety

Anxiety referred to feelings of fear, worry, and unease caused by external or internal potential threats (Grupe & Nitschke, 2013). Test anxiety was a situation-specific anxiety experienced in evaluative situations (Putwain, 2008), and generally consisted of two components: cognitive components, such as worry and test-irrelevant thinking, and affective components, such as emotionality and bodily symptoms (Zeidner, 1998). A great number of studies were devoted to exploring antecedents relate to test anxiety. Studies showed that anxiety has been found to be associated not only with negative attitude but also poor academic outcomes in the subject and low
Computer-based test

Computer-based test (CBT) was generally defined as an integrated procedure in which language performance is elicited and assessed by computers (Noijons, 1994). Being considered as an evolutionary step to future testing mode, CBT has been the research focus of educators and researchers. Some researchers investigated the technical aspects of CBT, ranging from item pool construction (He & Reckase, 2013) to comparison of different item-selection methods (Finkelman, Kim, Weissman & Cook, 2014; He, Diao & Hauser, 2014). Some researches focused on the correlation analysis, such as positive correlation between attitude and performance (Stricker & Attali, 2010), negative correlation between test anxiety and performance (Ortner & Caspers, 2011; Ortner, Weißkopf, & Koch, 2014; Lu, Hu, Gao, & Kinshuk, 2016), or no significant correlation between them (Shen et al., 2010).

From a brief review of the literature on test anxiety in CBT, it was quite obvious that so far the conducted studies generally focused on test administration, theoretical models, test reliability and validity, and the test takers’ general attitude. However, few studies have discussed test anxiety of the examinees in a specific way, such as the test anxiety differences of gender and at different language proficiency levels, and the test anxiety sources in computer-based testing environment.

Research design

In this paper, we first used a pilot study involving 61 students to testify and qualify the reliability of the questionnaire scale in the mid-semester and then at the end of the semester we conducted the experiment to investigate test anxiety of Chinese college learners of English as a foreign language in CBSET and made probation of the test anxiety causes.

Research questions

The main purpose of this paper was to explore specific test anxiety demonstration in CBSET, test anxiety differences in terms of gender and language abilities and the antecedents of test anxiety in CBSET. Thus, answers to the following questions would be discussed in the study.

- What’re the representations of test anxiety in CBSET?
- What’re the anxiety differences of genders and different language proficiency levels in CBSET?
- What’s the correlation between test anxiety and test performance in CBSET?
- What’re the test anxiety causes in CBSET?

Research instrument

The research was basically a quantitative study, including a computer-based spoken English test and a subsequent Likert-type questionnaire to measure students’ test anxiety in CBSET. SPSS 17.0 (Statistical Package for Social Science 17.0 version) was employed in the study. The questionnaire consisted of two parts: Part A was about students’ attitude toward spoken English and self-evaluation about their speaking abilities, and Part B was concerned with the scale to measure students’ anxiety in computer-based spoken English testing environment.

Since this paper attempted to pinpoint test anxiety under computer-based settings, Test Anxiety Scale constructed by Sarason (1978) and Attitude towards Computerized Assessment Scale by Smith (2003) were adopted in the study and synthesized as the new Test Anxiety Scale in CBSET (TACBSET). There were 26 items in total, of which 17 items in the scale were concerned with test anxiety scale, covering the following aspects: fear of negative evaluation, communication apprehension, test anxiety which further contained worry and emotionality. 9 items were related to attitude toward computerized anxiety scale, covering: attitude toward CBSET, practice in CBSET and anxiety in CBSET. Each item was to answer on a 5-point Likert scale, including “strongly disagree,” “disagree,” “uncertain,” “agree” and “strongly agree.”
In order to ensure the reliability of the questionnaire, pilot study has been conducted to 61 freshmen who were from the same academic background. The data of 61 students who took part in the first pilot study was analyzed through SPSS 17.0 to assess the quality of the questionnaire. Statistical analysis showed that the Cronbach reliability for the Test Anxiety Scale in CBSET was .892.

The result of the pilot study showed no problem for the participants in understanding the contents of the questionnaires and the reliability of the questionnaire was 0.892, which was satisfactory and safe enough to be adopted.

Research subjects

The participants in this study were 330 freshmen in a key university and they all had experience in taking computer-based spoken English test. The average age of the participants was 18, ranging from 16 to 20. Most of the subjects have been studying English as a second language for about 6 to 7 years. The final sample consisted of 229 students of 180 being males and 49 being females, after one student was rejected because his responses have missed values. In order to further discriminate the test anxiety differences, students at different language levels were chosen to take part in the test. The placements of their language abilities were based on their English scores in Chinese College Entrance Examination. 79, 90 and 60 were from advanced, medium and low level group respectively.

Research procedure

In computer-based spoken English testing environment, there were two supervisors, an English teacher and a technician. As students sit down to take the test, the Test Administration Module first asked them to check the machine to make sure the volume of their headsets were adequate to hear satisfactorily whatever instructions or questions to be presented and also to verify their voice being recorded properly in the designated storage area. Once the volume and recording checks completed, test items or questions were administered to the examinees in the order and manner specified by the computer programs automatically. During the administration of the test, the examinees generally did not have to perform any tasks other than speaking into their microphones and clicking on buttons to advance through the test.

In order to precisely capture test takers’ feelings, test administrators subsequently delivered questionnaires to test takers after they had taken computer-based spoken English test. All the subjects were instructed to fill in the questionnaire with their true opinions and feelings in the test and there was no right or wrong answer. The whole process lasted for about 12 minutes.

Results

After the collection and analysis of the data from computer-based spoken English test through SPSS 17.0, each of the 229 students’ scores could be obtained.

Test anxiety statistics in CBSET

In the new scale, 17 items were related to Test Anxiety (TA) and the score of each item ranged from 1 to 5 points, thus, the potential score of each student’s anxiety should range from 17 to 85 points. The 17 items were composed of three aspects: fear of negative evaluation (item 5 and 17), which was defined as the apprehension about other’s evaluations, avoidance of evaluative situation, and the expectation that others would evaluate one negatively (Wastson & Friend, 1969), communication apprehension (item 1, 14 and 16), which was a type of shyness characterized by fear of or anxiety about communication with people (McCroskey, 1977), and test anxiety referred to a type of performance anxiety stemming from a fear of failure (Sarason, 1980), which was in turn composed of two factors: worry (item 2, 3, 10, 11, 13 and 15) and emotionality (item 4, 6-9 and 12). Attitude towards Computerized Assessment (ATCA) consisted of 9 items, and the potential score of each student’s anxiety ranged from 9 to 45. This part was also composed of three factors: attitude toward CBSET (item 18-21), practice in CBSET (item 22 and 23) and anxiety in CBSET (item 24-26). The scale of Test Anxiety and the scale of Attitude towards Computerized Assessment finally made up the scale of Test Anxiety in
Computer-Based Spoken English Test (TACBSET). Thus, total speaking test anxiety score of each student ranged from 26 to 130 points.

<table>
<thead>
<tr>
<th>Table 1. Descriptive analysis of test anxiety in CBSET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>TACBSET</td>
</tr>
<tr>
<td>TA</td>
</tr>
<tr>
<td>ATCA</td>
</tr>
</tbody>
</table>

SPSS 17.0 has processed 229 valid individual data and it was found out that students’ speaking test anxiety registered between 38 (minimum) and 119 (maximum), with a range of 81. From Table 1, it could be seen that the general mean score of test anxiety in CBSET was 2.885, which was slightly below the average point 3. It showed that test takers were not as stressful as assumed in the CBSET. It was discovered that 48.5% of test takers were below the mean, 4.7% were on the mean and 46.8% were above the mean. The mean score of TA was 3.032, a little higher than the average level while the mean score of ATCA was 2.606, largely lower than the average point. Comparing the mean scores of TA and ATCA, we could find that participants perceived more anxiety on spoken English test than the anxiety caused by computers. Based on the general analysis of test takers’ test anxiety in CBSET, we have obtained a general view of test takers’ anxiety in CBSET. It was necessary to make a further exploration on the specific and detailed description of test anxiety.

<table>
<thead>
<tr>
<th>Table 2. Detail analysis of test anxiety in CBSET</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>TACBSET</td>
</tr>
<tr>
<td>TA</td>
</tr>
<tr>
<td>CA</td>
</tr>
<tr>
<td>TA1</td>
</tr>
<tr>
<td>w</td>
</tr>
<tr>
<td>e</td>
</tr>
<tr>
<td>sum</td>
</tr>
<tr>
<td>ATCA</td>
</tr>
<tr>
<td>A1</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>A2</td>
</tr>
</tbody>
</table>

Note. FNE: fear of negative evaluation (item 5 and 17); CA: communication apprehension (1, 14, 16); TA1: test anxiety (w: worry item 2, 3, 10, 11 and 13; e: emotionality 4, 6-9, 12); A1: attitude toward CBSET (item 18-21); P: practice in CBSET (item 22 and 23); A2: anxiety in CBSET (item 24-26).

Table 2 showed that the mean score for fear of negative evaluation was 2.335, 3.097 for communication apprehension, 3.117 for test anxiety with the component worry of 3.073 and component emotionality of 3.158. The scores of test anxiety and communication apprehension were both over the average point 3. It showed that among the three components, test anxiety occupied much portion, followed by communication apprehension and fear of negative evaluation. The high value of test anxiety suggested students were afraid of participating in the evalulative activities and were less confident of their communicative language abilities or were lack of communicative language abilities. As for the three components of attitude toward computer-based spoken English test, scores were 2.84, 2.535 and 2.66, respectively, all below the average point 3, which indicated that students did not feel much anxiety about the adoption of computers in spoken English test and generally accepted the new test format.

Anxiety differences in gender

Gender was commonly stated to have an effect upon the development and manifestation of anxiety in assessment situations. Men and women have been hypothesized to interpret and respond to evaluative situations in a differential manner (Arch, 1987).

The gender difference in test anxiety in CBSET could be detected via Independent Sample t-test. From Table 3, we could see that although the anxiety level of female students ($M = 2.962$) was higher than the one of male students ($M = 2.838$), the mean difference was not statistically significant ($t = 1.810$) (see Appendix1). In other words, though female students perceived more anxiety than male students, their anxiety degrees were close and the test anxiety differences were not significant. A deep look found that female students felt much anxiety on spoken English test with the mean anxiety value of 3.028, compared with the mean value of male students’ 2.981. With regard to the attitude toward CBSET, both females and males demonstrated much lower anxiety than spoken English test anxiety, with values of 2.839 and 2.569, respectively. To be concluded, the findings
showed no matter to spoken English anxiety and attitude toward CBSET, female students consistently perceived more anxiety than male students did and that both sexes had comparatively more anxiety on spoken English test.

<table>
<thead>
<tr>
<th>Table 3. Test anxiety differences in gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>TACBSET Female</td>
</tr>
<tr>
<td>TACBSET Male</td>
</tr>
<tr>
<td>TA Female</td>
</tr>
<tr>
<td>TA Male</td>
</tr>
<tr>
<td>ATCA Female</td>
</tr>
<tr>
<td>ATCA Male</td>
</tr>
</tbody>
</table>

Anxiety differences at different language proficiency levels

The number of subjects in this study was 229, of which comprised 79 students in advanced group, 90 students in medium group and 60 students in lower group.

<table>
<thead>
<tr>
<th>Table 4. Test anxiety differences at different language proficiency levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>TA</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ATCA</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>TACBSET</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The descriptive analysis of test anxiety scores in CBSET for different language proficiency levels indicated that the differences between any of the two levels were significant, that was to say, the differences among advanced, medium and lower language levels were distinguished. The test anxiety scores for each group had an inverse relationship with the group language proficiency levels, namely, the group with higher language levels achieved lower test anxiety score and vice versa. The mean scores for advanced, moderate and low language level groups were 2.77, 2.98 and 3.08, respectively (as shown in Table 4).

Through the observation of the values of range and the standard deviation, it could be seen that there existed a much significant anxiety difference in advanced language level group with the range of 81 and standard deviation .5165 and less significant anxiety difference in lower language level group with the range of 36 and standarddeviation.2654. The results suggested that test anxiety in advanced language levels appeared in an extreme way, that was, of the students with high language proficiency levels, some took CBSET with low anxiety while others with high anxiety. Perhaps some students with high language level were confident enough in their language abilities and took the examination easily, while some students were still anxious about their performance, probably, because they had much self-negative evaluations or they were less confident or they were afraid of failure, which imposed much stress on them, even though they had strong language abilities.

With comparison of scores of TAS and ATCAS, it could be found the most test anxiety perceived by students with different language levels were from the spoken English test rather than the new test style with adoption of computers. By checking the mean scores of spoken English test anxiety for each group, it was obvious that the scores of medium and lower level groups were over the average 3.0, with the score of 3.14 and 3.27, respectively, and the score of advanced group was 2.91, below the average one. It suggested that students with high language abilities were less stressful than those with lower ones, while taking the spoken English test. Besides, it could be seen that the mean scores of students’ attitude toward computer-based English test also had a positive relationship with students’ language abilities. In another words, students with higher language abilities hold positive attitude towards the new test style with computers.
Analysis of the effect of test anxiety on test performance

In the previous parts, we discussed examinees’ test anxiety in a micro and minor way and found males and females, no matter they were from high, medium or lower language proficiency levels, all perceived much anxiety portion in the respect of spoken English test anxiety and less anxiety on the adoption of computers in spoken English test. Since the major cause of anxiety was from spoken English test anxiety, it was necessary to clarify the relationship between the anxiety and test performance to see whether test anxiety had an adverse effect on test performance.

| Table 5. Correlation analysis between test anxiety and test performance (TP) |
|--------------------------------------------------|------------------|-----------------|-----------------|
| TP                  | TA               | ATCA            | TACBSET         |
| Pearson correlation | -166             | -058            | -153            |
| Sig. (2-tailed)     | .012             | .382            | .021            |
| N                   | 229              | 229             | 229             |

Note. *p < .05.

As shown in Table 5, it could be seen that the test anxiety in CBSET had a significantly negative correlation with the test performance with r of -.153. It suggested that students with higher test anxiety in the test would achieve lower test scores and students with lower test anxiety in the test would achieve higher test scores. To explore further, we could find that spoken English test anxiety also had a significantly negative correlation with test performance with r of -.166, while attitude toward computer-based spoken English test had a weak negative relationship with the test performance.

Test anxiety sources

As we have discussed the adverse effect of test anxiety in CBSET on test performance, it was essential and necessary to do a further exploration of the test anxiety sources so as to have a full understanding and made preparations to reduce the anxiety. In the questionnaire, students were asked to arrange the anxiety source of low spoken language abilities, inadaptability in computer-based spoken English test, anxiety from the evaluation process and lack of speaking techniques in the CBSET.

| Table 6. Major anxiety source |
|-------------------------------|-----------------|-----------------|-----------------|
| Valid                         | Frequency       | Percentage      | Valid percent   | Cumulative percent |
| AS1                           | 136             | 59.5            | 59.5            | 59.5              |
| AS2                           | 7               | 3               | 3               | 62.5              |
| AS3                           | 29              | 12.6            | 12.6            | 75.1              |
| AS4                           | 57              | 24.9            | 24.9            | 100               |

Note. AS1: low spoken language abilities; AS2: inadaptability in computer-based spoken English test; AS3: anxiety from the evaluation process; AS4: lack of speaking techniques.

In Table 6, among 229 research subjects, 136 students showed that their major anxiety was from low spoken language abilities, taking up 59.5% of the total. 7 students told that their major anxiety source was the inadaptability in CBSET, only occupying 3%. 29 of 12.6% students chose the anxiety of taking test as the major one, and 57 students believed lack of speaking techniques constructed the major anxiety, accounting for 24.9% of the total.

Among the 79 students from advanced language level (as shown in Table 7), 42 students (53.2%) showed that their major anxiety resulted from lower spoken language abilities. Only 3 students (3.8%) said their major anxiety was from the inadaptability of CBSET. 13 students (16.5%) expressed their major anxiety from the test itself and 21 students (26.6%) chose lack of speaking techniques as the major cause of their anxiety. In the medium level group, 50 students (55.6%) chose the low spoken language abilities as the major anxiety source. 3 students (3.3%) held the inadaptability in CBSET, 12 students (13.3%) believed the test made them feel anxious, and 25 students (27.8%) believed lack of speaking techniques was the major source of anxiety. Of the lower level group, 44 students (73.3%) expressed that their major anxiety source was from low spoken language abilities. only one student (1.7%) showed the inadaptability in CBSET. Four students (6.7%) chose the test itself and 11 students (18.3%) believed lack of speaking techniques was the major anxiety source.
Table 7. Major anxiety source for students at different language levels

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>AS1</td>
<td>Advanced</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>AS2</td>
<td>Advanced</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AS3</td>
<td>Advanced</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>AS4</td>
<td>Advanced</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Lower</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 8. Correlation analysis of self-evaluation on spoken English and task importance

<table>
<thead>
<tr>
<th></th>
<th>Self-evaluation on spoken English</th>
<th>TACBSET</th>
<th>TP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-evaluation on spoken English</td>
<td>Pearson correlation</td>
<td>-0.238**</td>
<td>0.477**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>229</td>
<td>229</td>
</tr>
<tr>
<td>Task importance</td>
<td>Pearson correlation</td>
<td>0.167*</td>
<td>-0.017</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>0.011</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>229</td>
<td>229</td>
</tr>
</tbody>
</table>

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

From the above table, it could be seen that self-evaluation on spoken English and task importance variables could influence the test anxiety in CBSET and test performance. Table 8 showed that self-evaluation on spoken English was significantly negatively correlated with the test anxiety in CBSET and significantly positively correlated with the perceived test performance, that is to say, students who favored in spoken English and with high self-evaluation of speaking abilities perceived lower test anxiety in CBSET and achieved higher test scores and vice versa. Besides, it still could be found that students’ perception of task importance did not have much influence on the test anxiety but had a significantly positively correlation with the test performance. That was, whether students valued the test or not, their reaction would not significantly affect the test anxiety in CBSET but their attitude could greatly influence their performance. Students who attached much more importance to the test would achieve higher scores and those who attached less importance to the test would get lower scores.

Discussion

Findings of the study indicated that students did not perceive or show much more anxiety about spoken English test based on computers as what we originally believed. During the test, students perceived less anxiety towards the adoption of computers in spoken English test than that generated from the self-efficacy of their speaking abilities and apprehension of failure in the test. The results were in agreement with some researches (Zhang, 2001; Rahimi & Zhang, 2016). In view of this finding, it could be explained that what students were really concerned about more were their poor speaking abilities and apprehension of failing in the spoken English test. With regard to the correlation between test anxiety and performance in CBSET, the results indicated that students’ test performance was significantly influenced by spoken English test anxiety (Chapell et al., 2005;
rather than the anxiety aroused by the adoption of computers in the test. The findings were in line with the results of previous studies (Ortner & Caspers, 2011; Trifoni & Shahini, 2011; Iroegbu, 2013). Though attitude toward computer-based spoken English test had a weak relationship with test performance, it was found that there existed a significantly negative relationship between spoken English test anxiety, which was consistent with the finding of Ortner and Caspers (2011). In other words, students with higher spoken English test anxiety would hold negative attitude toward CBSET and vice versa. It has a strong negative correlation with spoken English test anxiety which has a strongly adverse effect on test performance. Thus, the attitude towards computer-based spoken English test could not be neglected but should be taken into account in test.

Concerning to anxiety differences of genders and different language proficiency levels in CBSET, the results appeared that gender test anxiety differences were not significant, though female students perceived a little more test anxiety than that of male students. As expected, the findings were consistent with those obtained in studies of Sam, Othman and Nordin (2005), and Kannan, Muthumanickam and Chandrasekaran (2012). This could be explained by the higher level of emotional response and perceptions of threat for females in testing environment. As to the question of test anxiety difference in different language proficiency level, the study showed that the higher language proficiency level students had, the less anxiety they would perceive. The findings were in agreement with research results of Kitano (2001). Students with higher language proficiency levels had more confidence in spoken English abilities and students with lower language proficiency levels had less confidence in spoken English abilities. Besides, no matter what language proficiency levels students have, their choices for major anxiety source were consistently the lower spoken language abilities and choices for minor anxiety source were the inadaptability in CBSET. Thus, it could be concluded that the result was another supplement to prove that adoption of computers in spoken English test did not construct much effect on test takers’ performance in CBSET.

The findings of the research can be applicable in both language test field and language teaching pedagogy. In consideration of the adverse effect of spoken English test anxiety on students’ test achievement in CBSET, it is necessary to enhance the awareness of spoken English test anxiety and take measures to reduce the anxiety. However, test takers’ feelings and emotions can be affected by many factors, external or internal ones. And these factors cannot be checked in an all-round way in the study. This paper just touched the tip of the iceberg of the computer-based spoken English test. Thus, more investigations and researches need to be done in the field of examinees’ anxiety and a further study of how to reduce test anxiety in CBSET will be conducted.

Conclusion

This paper has provided evidence for the discussion of test anxiety of spoken English in computer-based testing environment. The research results showed that the major anxiety perceived by test takers was the spoken English test anxiety. The results told that of the test anxiety in CBSET, spoken English test anxiety and communication apprehension were the two major anxiety components. Although the current study was insufficient and cannot cover all the aspects of the test anxiety issues in CBSET, its results and conclusions were expected to increase students’ awareness of spoken English test and alter peoples’ traditional ideas against the adoption of computers in spoken English test. It was also expected that this paper would usher in more researches on the test anxiety and relief measures of test anxiety in computer-based spoken English test.

Acknowledgements

Key Project for Excellent Young Talents in Colleges and Universities in Anhui Province (gxyqZD2016385).

References


Appendix 1

Independent Samples Test

Table 9. Levene’s test for equality of variances (t-test for equality of means, 95% confidence interval of the difference)

<table>
<thead>
<tr>
<th>Test anxiety</th>
<th>F</th>
<th>Sig.</th>
<th>t</th>
<th>df</th>
<th>Sig.(2-tailed)</th>
<th>Mean difference</th>
<th>Std. error difference</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal variance assumed</td>
<td>2.427</td>
<td>.121</td>
<td>1.810</td>
<td>227</td>
<td>.072</td>
<td>3.226</td>
<td>1.782</td>
<td>-.285</td>
<td>6.738</td>
</tr>
<tr>
<td>Equal variance not assumed</td>
<td>1.647</td>
<td>67.887</td>
<td>.104</td>
<td></td>
<td></td>
<td>3.226</td>
<td>1.959</td>
<td>-.684</td>
<td>7.136</td>
</tr>
</tbody>
</table>

Appendix 2

Questionnaire for Test Anxiety in Computer-based Spoken English Test

Student No.: ________ Age: ________ Sex: ________

Part A:
Direction: Read the following statements and make a √ in the box next to the statement that fits your ideas.
1. Your attitude toward learning spoken English: [ ] like; [ ] uncertain; [ ] dislike
2. Self-evaluation of speaking abilities:
   [ ] Good. I can speak English very fluently.
   [ ] Medium. Although I find it a bit difficult in speaking English, I can basically express my ideas in daily conversations.
   [ ] Bad. I can only speak a few and simple words in communication.
3. Importance perception of this test: [ ] value; [ ] uncertain; [ ] not value

Part B:
Direction: Read each items below to see if it reflects your experience in test taking. If it does, place a check mark on the line next to the number of the statement. Check as many as seem fitting. Be honest with yourself.
1=totally disagree, 2=disagree, 3=uncertain, 4=agree, 5=totally agree
1. Before taking a test, I thought other students’ speaking abilities are higher than mine.
2. I have an uneasy, upset feeling before taking the examination.
4. In the test, I found myself thinking of things unrelated to the test.
5. In the test, I found myself thinking of the consequence of failing.
6. In the test, I felt my heart beating very fast.
7. After taking a test, I always feel I could have done better than I actually did.
8. I usually get depressed after taking a test.
9. During the examination, I frequently get so nervous that I forget facts I really know.
10. The harder I work at taking the test or studying for one, the more confused I get.
11. I wish examinations did not bother me so much.
12. I think I could do much better on tests if I could take them alone and not feel pressured by time limits.
13. I really don't see why some people get so upset about tests.
14. Thoughts of doing poorly interfere with my performance on tests.
15. Even when I'm well prepared for a test, I feel very anxious about it.
16. I am unconfident of myself when I speak in English.
17. I’m afraid of being looked down if I fail in the test.
18. I hope I can take computer-based spoken English test.
19. I feel uneasy when I take computer-based test.
20. I think computer-based spoken English test can assess my real speaking abilities.
21. I have confidence in taking computer-based test.
22. Practice in computer-based test can improve test performance.
23. I think I can be accustomed to computer-based test if I have more practice.
24. I’m afraid that lacking of computer experience or skills would influence my test performance.
26. Speaking English at the same would interfere with my thoughts.
**Integrating Dynamic Mathematics Software into Cooperative Learning Environments in Mathematics**

Yılmaz Zengin\(^1\) and Enver Tatar\(^2\)

\(^1\)Department of Mathematics and Science Education, Dicle University, Turkey // \(^2\)Department of Mathematics and Science Education, Atatürk University, Turkey // yilmazzengin@outlook.com // entatar@gmail.com

\(^*\)Corresponding author

(Submitted June 8, 2016; Revised August 19, 2016; Accepted October 13, 2016)

**ABSTRACT**

The aim of this study was to evaluate the implementation of the cooperative learning model supported with dynamic mathematics software (DMS), that is a reflection of constructivist learning theory in the classroom environment, in the teaching of mathematics. For this purpose, a workshop was conducted with the volunteer teachers on the implementation of the cooperative learning model supported with DMS. Dynamic materials and worksheets suitable for quadratic functions and sequences topics were developed. The effect of implementing the cooperative learning model supported with DMS in the teaching of the quadratic functions and sequences topics on student performance as well as students’ views about the model were examined. The study was carried out using an embedded design. The study group consisted of 61 high school students. A quadratic functions knowledge test, a sequences knowledge test, and an open-ended questionnaire were used as data collection tools. The Mann–Whitney test and dependent t-test were used for the analysis of quantitative data, while content and descriptive analyses were used for the analysis of qualitative data. As a result of analysis of the data, it was found that the model had a positive effect on student achievement. Moreover, the following students’ views were identified: the model enabled better understanding, it visualized and concretized the course, and it created a pleasant and enjoyable learning environment.

**Keywords**

Dynamic mathematics software, Cooperative learning, GeoGebra, Mathematics learning and teaching

**Introduction**

Social changes have triggered rapid development. Information and communication technologies (ICTs) have affected human life and have resulted in new opportunities to emerge and new information to be created. These innovations require perspectives and expectations towards mathematics, the use of mathematics in different ways and a review of mathematics instruction. In line with these changes and developments, new problems have occurred in our life each passing day and individuals who have realized the importance and value of mathematics and who have developed their power of thinking are needed more than ever (Turkish Ministry of National Education [TMoNE], 2013). However, as mathematics is more abstract and difficult (Herzig, 2002) when compared to other disciplines (Dienes, 1971; Frenkel, 2013; Sarama & Clements, 2009), it results in difficulties in the teaching and learning of mathematics (Yenilmez & Avçu, 2009).

As most mathematics concepts are abstract, they cause difficulties in the teaching and learning processes. Functions are one of the most important concepts which are difficult to learn, resulting in misconceptions (Ural, 2006) and play both a central and connective role among mathematics topics (Selden & Selden, 1992). As a result of the close relationship between functions and mathematical thinking, the active use of functional thinking for solving problems in different disciplines (Bayazit & Aksoy, 2013), the important role of learning functions and graphs in understanding mathematics (Kutluca & Baki, 2013) and the difficulties while learning the topic are increasing more and more. Quadratic functions and graphs are special and important states of functions and they are the basic principles of functions for secondary school students (Even, 1990). However, it is known that students have difficulty with quadratic functions and graphs just like with functions (Kutluca & Baki, 2009; Sajka, 2003; Tatar, Okur, & Tuna, 2008; Zazkis, Liljedahl, & Gadowsky, 2003). After the functions topic is taught in the 9th grade in the secondary school mathematics curriculum, students again encounter quadratic functions and graphs in the 10th grade in Turkey. Similarly, in 11th grade, there are topics which mainly involve functions. The sequences topic which is linked to the functions topic is one of the main topics in which students have difficulty in understanding in the 11th grade in Turkey (Akgün & Duru, 2007; Durmuş, 2004; Tatar, Okur, & Tuna, 2008).

Functions, quadratic functions and graphs, sequences, numbers and algebra are generally among the difficult topics to learn. Students continue to meet these topics in the numbers and algebra learning domain of the curriculum in the mathematics courses taught in their undergraduate studies, particularly relating to the
knowledge of functions. It is important that these concepts which are used as basic ideas in mathematics courses like advanced analysis and differential equations, should be learned (Bayazıt, 2010). Therefore, it is considered that research studies on these topics which are included in the learning domain of numbers and algebra would be useful to students, mathematics teachers, and mathematics educators.

Boosting mathematical topics which involve abstract concepts with ICT can enable mathematics to be concretized and help concepts to be learned more easily (Baki, 2002). Moreover, students who are actively engaged in the problem-solving process with the use of ICTs, experience different representations of concepts rather than repeated calculations, and work on real mathematics problems (TMoNE, 2013). Open source dynamic mathematics software (DMS) GeoGebra is a freely available and easy-to-use ICT (Diković, 2009; Hohenwarter & Preiner, 2007).

Examining and utilizing ICTs within the scope of suitable learning approaches is as important as choosing the most appropriate tool. The role of technology is not to transfer knowledge but to enable a learner to construct knowledge and make sense of their experiences based on the constructivist learning theory (Tezci & Perkmen, 2013). The cooperative learning model is considered as an important tool which promotes the reflection of constructivist learning theories that regard learning by discovery and learning as a social activity in the classroom environment (De Lisi & Golbeck, 1999). There are many methods used in the cooperative learning model. One of them is the Student Teams Achievement Division (STAD) and it is implemented in many majors such as mathematics, science, social studies and industrial arts at all levels from primary education to higher education (Slavin, 1994). Moreover, because the STAD motivates students, it teaches them to help each other (Slavin, 1987) and as it is practiced easily (Slavin, 1995), it was preferred in this study as GeoGebra. GeoGebra is a suitable ICT software that uses the cooperative learning model based on the constructivist learning theory, particularly for teaching of difficult concepts like quadratic functions and sequences.

The DMS has a positive effect on student achievement and motivation in mathematics. When teachers use this software in mathematics classrooms students increased attention and interest engender better understanding and facilitate conceptual learning (Choi, 2010; Tatar & Zengin, 2016). When the literature about using DMS in learning and teaching of difficult mathematics concepts is examined, it is found that DMS provides positive effects such as an increase in achievement (Ayvaz Reis & Özdemir, 2010; Saha, Ayub, & Tarmizi, 2010), an interesting, visual and concrete learning environment (Tatar & Zengin, 2016), and opportunities for problem-solving. Considering the positive findings about how to integrate this dynamic software in the learning environment, it is very important for educators to evaluate these positive capabilities of DMS. On the other hand, one of the important skills in the 21st century is the inculcation of cooperative skills. Cooperative learning facilitates the development of improved inter-group relations, social acceptance, increased self-esteem, and mathematics achievement. The other contributions of this method are that students improve their problem-solving abilities and the ability to integrate specific knowledge and skills (Slavin, 1995). In this regard, DMS GeoGebra was embedded in cooperative learning in this study for integrating the strengths of both the software and the cooperative learning model. The cooperative learning model supported with DMS was implemented in difficult concepts in mathematics. When the literature is examined, it is found that students have difficulty in understanding about quadratic functions (Kutluca & Baki, 2009; Sajka, 2003; Tatar, Okur, & Tuna, 2008; Zazkis, Liljedahl, & Gadowsky, 2003) and about sequences (Akgün & Duru, 2007; Durmuş 2004; Tatar, Okur, & Tuna, 2008). There are many separate studies about the use of DMS and cooperative learning in mathematics. However, the use of both quadratic functions and sequences in high school level is not found in the literature. The significance of this research is that it may provide a good alternative model for teaching difficult mathematics concepts such as quadratic functions and sequences in high school mathematics. In addition, the other significance of this research is that it may contribute to integrating DMS in the learning environment using the cooperative approach. Therefore, a significant attention to the effectiveness of the model at the high school level of mathematics teaching is crucial. In this context, the research questions are as follows:

- What is the effect of the cooperative learning model supported with DMS on student achievement in mathematics?
- What are the views of students on the use of the cooperative learning model supported with DMS in mathematics?

**Method**

Both qualitative and quantitative data were gathered and analyzed concurrently or sequentially throughout the research process using an embedded design. One of the purposes of this design is that the qualitative data obtained must support the quantitative data or vice versa. Moreover, this design is suitable for interpreting
different research problems with different types of data. The advantage of this design is that it integrates the benefits of both quantitative and qualitative data. The weakness of this design is that the two sets of data may be difficult to compare because each data set addresses different research questions. Furthermore, the data collection is very labor intensive for a single researcher (Creswell, 2012). In the study, the effect of the implementation of the cooperative learning model supported with DMS in the quadratic functions and sequences topics on student achievement was evaluated using the quantitative data and students' views about the model. The embedded design was used in this research study because it provided opportunities to collect data for the different problems of the study using different data collection procedures. There were two experimental groups, one each for sequences and quadratic functions, as well as a control group in this study. The cooperative learning model supported with DMS was implemented with the experimental group for the sequences topics for three weeks. The control group was instructed in sequences only in a traditional method for three weeks. The experimental group for the quadratic functions topic was instructed in the cooperative learning model supported with DMS for two and a half weeks.

Study group

The study group consisted of 61 high school students in Turkey of whom 27 were male and 34 were female students. Participants’ ages ranged from 16 to 17 years. The experimental group for the sequences topic, the experimental group for quadratic functions, and the control group consisted of 19, 25, and 17 high school students, respectively. The participants were chosen using the convenience sampling method. This sampling method is preferred because of students’ convenient accessibility and proximity to the researchers (Yıldırım & Şimşek, 2011).

Data collection tools

Using the data obtained with different methods increases the reliability and validity of the results obtained (Yıldırım & Şimşek, 2011). Therefore, both qualitative and quantitative data collection tools were used in the research study. In the research, the Sequences Knowledge Test (SKT), the Quadratic Functions Knowledge Test (QFKT), and an open-ended questionnaire were used. The data collection tools used in the study are discussed in detail below.

Knowledge test

The SKT and QFKT were administered as pre-test and post-test. While developing the test, the TMoNE (2011) and TMoNE (2013) were referred to by the researchers. Four faculty members from the mathematics education department and three mathematics teachers were asked to review the test for the content validity of the tests that were developed as well as the accuracy of the items. Necessary corrections were made based on the feedback received and the test was finalized. A holistic evaluation form was developed based on the TMoNE (2011) and the evaluation of the test at the end of the implementation was carried out using this form.

Open-ended questionnaire

After the model was implemented in the classroom, a questionnaire was administered to the students with the intention of examining its feasibility. Before the questionnaire was prepared, the relevant literature was reviewed, the views of two experts from mathematics education were considered and draft forms were developed. During the process of piloting, the questionnaire was distributed to 62 students. The questionnaire gathered at the end of the piloting was examined, necessary corrections were made and the questionnaire was finalized to be used in the research study. The addition of the following question to the questionnaire can be given as an example of the corrections that were made: “Do you have any other opinions and suggestions about the implementation process?” One of the questions included in the open-ended questionnaire was: “What effects did the DMS and the cooperative learning model have on you?”
Implementation process

A practice teacher participated in the workshop and he volunteered to implement the model in his class. A guide book which could be used as a supplementary resource about DMS was prepared by the researchers was distributed to the teachers. The workshop consisted of 18 sessions; each session lasted 50 minutes and was held on four days. The aim of the first 11 sessions of the workshop was to introduce the software. In the last seven sessions, information about the implementation of the cooperative learning model supported with DMS was given.

![Diagram showing the implementation process]

Every four students used one computer and two worksheets in order to enforce teammates to work together (Slavin, 1994). The researchers downloaded the materials which were planned to be used that week before the lesson. Base scores, determined based on students’ final grades in the previous semester, were used for determining initial scores. Teams were composed based on these base scores and the chart developed by Slavin (1994). After the teams were identified, the volunteer teacher introduced the model in his class. How the model helped to teach difficult mathematical concepts more effectively, is presented in detail in Figure 1.

Worksheets and dynamic materials

Ten dynamic materials and six worksheets were developed about the topic on sequences. For quadratic functions, six dynamic materials and four worksheets were developed. While developing the dynamic materials and worksheets, the website http://www.geogebra.org/ and Hohenwarter and Hohenwarter (2012), TMoNE (2011), TMoNE (2013), Şişman, Lökçü, Oğuz and Atak (2013) were referred to. After the worksheets and dynamic
materials were developed by the researchers, two mathematics educators and two mathematics teachers were asked to review them. Based on the feedback from the experts, the worksheets and dynamic materials were revised. Then, the materials were reviewed in the workshop with the teacher who volunteered to implement the model in his class. At the weekend meetings some corrections were made. Some examples from the worksheets which were used with the dynamic materials are given the Appendix.

Data analysis

Both qualitative and quantitative data were used together in the research. Quantitative data were obtained from the SKT and QFKT results. As the study group consisted of less than 50 people, the Shapiro–Wilk test was used to check whether the quantitative data obtained from the measurement results came from a normally distributed population (Büyüköztürk, 2011). The results of the Shapiro–Wilk test were as follows: $p_{\text{pretest-Experimental-I}} < .05; p_{\text{posttest-Experimental-I}} < .05; p_{\text{posttest-Control Group}} < .05; p_{\text{posttest-Experimental-II}} > .05; p_{\text{posttest-Experimental-II}} > .05).$ In addition, the Q-Q, box and whisker plots, detrended normality plot, kurtosis and skewness values were analysed to determine whether the measurement results exhibited a normal distribution or not (Field, 2009). Measurement results of both groups have to exhibit a normal distribution to choose parametric tests. The Mann–Whitney test was used for analyzing the quantitative data. The SPSS 18.0 program was used to for the analysis of the quantitative data obtained from the research. The value of 0.05 which is frequently used in education research studies is considered as the minimum significance level.

The qualitative data obtained from the open-ended questionnaires were analysed using content and descriptive analyses methods. The data obtained in the descriptive analysis were examined in detail using content analysis. Summarized data in the descriptive analyses were coded in order to examine the data in depth. Codes were used for different categories. The researchers consulted expert academicians to finalize the coding process and constitution categories (Yıldırım & Şimşek, 2011). Moreover, the number of students was taken as the frequency in the data analysis. “F” stood for the frequency of occurrence of the codes. Volunteer students for the sequences were codes as S1, S2, S3, …, S19. In addition, volunteer students for the quadratic functions topic were coded as Q1, Q2, Q3, …, Q25. The findings obtained as a result of the analysis of the data obtained from students included descriptions using these codes.

Findings

The Mann–Whitney test was used to examine whether there was a significant difference between the students in the experimental group and the control group based on their pre-test and post-test scores of the SKT. The students’ pre-test and post-test results are presented in Table 1 and Table 2 respectively.

<table>
<thead>
<tr>
<th>Group</th>
<th>$n$</th>
<th>Mean rank</th>
<th>Sum of rank</th>
<th>$U$</th>
<th>$z$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental-I</td>
<td>19</td>
<td>18.92</td>
<td>359.50</td>
<td>153.50</td>
<td>-.257</td>
<td>.797</td>
</tr>
<tr>
<td>Control</td>
<td>17</td>
<td>18.03</td>
<td>306.50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to Table 1, the Mann-Whitney test indicated that there was no significant difference between the experimental-group$_5$ (Mdn = 14) and the control group (Mdn = 16) based on their pre-test scores, $U = 153.50, p > .05, r = -.04.$ In terms of these results, both the experimental-group$_5$ and the control group were equivalent based on their pre-test scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>$n$</th>
<th>Mean rank</th>
<th>Sum of rank</th>
<th>$U$</th>
<th>$z$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental-I</td>
<td>19</td>
<td>22.26</td>
<td>423.00</td>
<td>90.00</td>
<td>-2.281</td>
<td>.023</td>
</tr>
<tr>
<td>Control</td>
<td>17</td>
<td>14.29</td>
<td>243.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to Table 2, the Mann-Whitney test indicated that the students of the experimental-group$_5$ (Mdn = 36) had significantly higher scores than the students of the control group (Mdn = 20), $U = 90.00, p < .05, r = -.38.$ Based on these findings, it can be stated that the cooperative learning model supported with DMS is an effective model to increase student achievement in the sequences topic.
The dependent t-test was used to examine whether there was a significant difference between the students’ scores in the pre-test and the post-test of the QFKT. The students’ test results are presented in Table 3.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>24</td>
<td>17.62</td>
<td>11.27</td>
<td>23</td>
<td>-2.30</td>
<td>.030</td>
</tr>
<tr>
<td>Post-test</td>
<td>24</td>
<td>20.79</td>
<td>11.25</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The analysis results given in Table 3 reveal that there was a significant difference in the scores for the post-test ($M = 20.79, SD = 11.25$) compared to the pre-test ($M = 17.62, SD = 11.27$), $t(23) = -2.30, p < .05, r = .43$. These results suggest that the cooperative learning model supported with DMS is an effective model to increase student achievement in quadratic functions.

As a result of the analysis of data obtained from the open-ended questionnaire of the 43 students who volunteered in the study after the implementation of the model, a total of four categories were identified: “contributions of the model,” “Obstacles to the implementation of the model,” “Materials,” and “Topics and lessons.” Table 4 presents the codes and categories of these codes as well as the frequency (f) of students involved.

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributions of the model</td>
<td>Enabling better understanding</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Visualizing the course</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Increasing retention</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Concretizing the course</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Pleasant and enjoyable learning environment</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Promoting interest and motivation</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Facilitating learning</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Having efficient lessons</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Learning from team work</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Learning the concepts via discussion</td>
<td>4</td>
</tr>
<tr>
<td>Obstacles to the implementation of the model</td>
<td>Taking time to get used to it</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Lack of computer skills</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Disagreements within the group</td>
<td>1</td>
</tr>
<tr>
<td>Materials</td>
<td>Suitability of materials</td>
<td>34</td>
</tr>
<tr>
<td>Topics and lessons</td>
<td>Implementing it in other courses</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Implementing it all the concepts of mathematics</td>
<td>3</td>
</tr>
</tbody>
</table>

When Table 4 is examined, for the framework of “contributions of the model,” students’ positive views regarding the cooperative learning model supported with DMS stand out. Students stated that as a result of the model, they understood the lesson better, the model visualized the course, it created a pleasant and enjoyable learning environment, and it increased retention. Moreover, they stated that the model did not have any difficulties, while their interest and motivation increased with the use of the model, and the lessons became more efficient.

A total of 15 of the students stated that the model promoted better understanding. Some students’ views were as follows:

“... We refreshed our knowledge. At the same time, it enabled us to understand the subjects better ....” (S1)

“... I understood a topic better for the first time.” (S2)

“I could not understand the topic that I want to learn in the classroom. But using the cooperative learning model and DMS together provided better understanding of difficult mathematics concepts.” (S8)

Fourteen of the students stated that the model enabled them to visualize the course. Two students’ views exemplifying this situation are:

“One of its advantages for the mathematics course is that it can be visualized with models ...” (S4)

“Its implementation in the mathematics course was good both for me and my friends because the mathematics course became more visual and concrete.” (S10)

Eight of the students stated that the model concretized the course. Two students’ views exemplifying this situation are:

“Since the model provides concretization our knowledge is more permanent” (Q6)
“In my opinion it is very effective, we see abstract mathematics concepts visually. We concretized the mathematics concepts in our team...” (Q14)

Thirteen students stated that the model increased retention and they did not have any difficulties. The following are some students’ views:

“We understood much better. We both loved mathematics and we learned that it was much simpler. Moreover, thanks to the cooperative learning model and dynamic mathematics, we remembered easily. Therefore, therefore the two combinations provided a positive learning environment for the sequences topic” (S 17)

“The use of both GeoGebra and cooperative learning approaches is a very effective way to increase retention.” (Q2)

“It is more permanent because of hands-on learning” (Q24)

“I think that this study did not pose any difficulties to us because this study made it much easier for us.” (S 10)

Seven students mentioned that the model created a pleasant and an enjoyable learning environment. Two students’ views are as follows:

“I believe that it has positive effects on me and my peers. It is a kind of enjoyable learning technique in which we listen and implement without getting bored and it really has a lot of benefits.” (S5)

“I realized that mathematics was pleasant, fun, and enjoyable.” (S11)

Two of the seven students who stated that the model increased their interest and motivation shared their views as follows:

“The combination of dynamic mathematics software and cooperative learning increased my interest and motivation, and this combination provided better understanding of quadratic functions.” (Q1)

“In fact, working in groups was very beneficial. As I refreshed my knowledge, my question repertoire expanded and my interest in mathematics increased.” (S3)

When Table 4 is examined regarding “obstacles to the implementation of the model”, it is found that the students had difficulties with the cooperative learning model supported with DMS and these difficulties usually resulted from disadvantages such as lack of computer skills, taking time to get used to the model, and disagreements within the group. Two students’ views can be given as examples of this category:

“At the beginning I did not understand anything but sometime later I began to understand.” (S2)

“At first we had difficulty in using the materials, but later we got used to them... Because they were easily understood, we liked them a lot.” (S5)

When Table 4 is examined regarding the “materials” category, most of the students stated that in the cooperative learning model supported with DMS, the materials used with the model were suitable. Two students’ views on this issue are given below:

“Dynamic materials about sequences brought different perspectives to the topic in order to learn that subject...” (S12)

“According to me, we learn better quadratic functions thanks to dynamic materials.” (Q4)

When Table 4 is examined regarding “topics and lessons”, students stated that they did not want the cooperative learning model supported with DMS to be implemented with only mathematics topics but with all other courses as well. The following view of S15 is an example:

“I wish we had such implementation with all courses. As we particularly have difficulties in numerical courses and sometimes we are afraid of our teachers, it would be much more productive if we worked with our peers.”

The feasibility of the cooperative learning model supported with DMS in classes was examined based on students’ views. It was revealed that the students understood the course much better with the model and it created a pleasant and an enjoyable learning environment. Moreover, the students suggested that the model increased retention and it affected their interest and motivation positively. In addition to this, they stated that the model did not have any difficulties, and it was found that the materials used with the model were instructive and they could be used in the classroom. In this context, when the views of Q1, Q2, S8, and S17 were examined, they stressed that the combination of the cooperative learning model and the dynamic mathematics software provided these positive contributions. However, the students mentioned that there could be difficulties with the implementation of the model because of some reasons such as disagreements within the group, taking time for them to get used to the model, and lack of computer skills.
Results and recommendations

In this study, the effect of the cooperative learning model supported with DMS on student achievement of difficult mathematical concepts such as sequences and quadratic functions was investigated. In addition, student views about this model were examined in detail. The analysis of quantitative data showed that this integrated model increases student achievement in the sequences and quadratic functions topics. This result is similar with previous researches that examined the use of cooperative learning in mathematics classrooms (e.g., Bernero, 2000; Kumar & Harizuka, 1998; Nichols & Miller, 1994; Stevens & Slavin, 1995; Ural, 2007; Vaughan, 2002). However, these studies generally focused at the elementary mathematics level. Similarly, this result is parallel with the previous researches which investigated the use of DMS in mathematics classrooms (Tatar & Zengin, 2016; Thambi & Eu, 2013; Zengin & Tatar, 2014; Zengin & Tatar, 2015). But these studies do not explain how to use DMS effectively in a cooperative learning environment that plays a very significant role in acquiring one of the important skills in the 21st century. Considering the quantitative results about sequences and quadratic functions, the cooperative learning model supported with DMS played significant role in increasing student achievement in these difficult mathematical topics. This model may contribute to students’ learning of difficult mathematical concepts from elementary level to university level.

It was found in other studies that students understood concepts better in the settings where DMS was used (Arbain & Shukor, 2015; Thambi & Eu, 2013; Zengin & Tatar, 2014; Zengin & Tatar, 2015) and their interest and motivation in the course increased (Choi, 2010; Green & Robinson, 2009; Zengin & Tatar, 2014). DMS help students achieve better understanding and conceptual learning (Tatar & Zengin, 2016). Similarly, it was revealed by other researchers that cooperative learning increased the motivation levels of students (Slavin, 1987; Nichols & Miller, 1994). It can be implied that because the model uses software and cooperative learning together, the results of this study are parallel to the results of other studies that have been conducted. It was found based on students’ views that the visualizations used in the model made the concepts clearer and an enjoyable learning environment was created with the model. The finding that the lessons which used DMS became more visual (Thambi & Eu, 2013; Diković, 2009) also coincide with the results of this study due to the software included in the model.

The model can enable students to understand concepts better and more easily by visualizing the concepts, creating a pleasant and an enjoyable learning environment, and increasing student motivation and interest and thus affecting students’ performance positively. It can be inferred that the cooperative learning model supported with DMS provides a good alternative learning model for understanding difficult mathematics concepts. Thus, teachers may have beneficial opportunities to use educational technology in mathematics classroom.

Based on the qualitative data obtained, students’ lack of computer skills must be addressed so that the model can be more applicable. It is important that students must be introduced to basic ICTs in computer courses and also they must be taught the content of the basic software specific to the field and how to use them in education. Accordingly, the content of computer courses can be updated. Moreover, it is observed that it will take some time for students to get used to the model and they can have disagreements within the group. It is suggested that teachers should implement the model and design the materials and worksheets considering these disadvantages. For this purpose, it is recommended that teachers should use relevant internet resources and also visit the official website of GeoGebra (www.geogebra.org) so that they can design the dynamic materials and worksheets themselves.

If researchers who would like to carry out studies on the model focus on how to eliminate the problem of time, disagreements within the group and lack of computer skills, they can make positive contributions to the model. Moreover, researchers can make contributions to the field by examining the effect of the model on students and teachers in other subjects.

Limitations

The main limitation of this study is about the quantitative part of the study. The quantitative part of study regarding quadratic functions was designed as a single group pretest-posttest. Additionally, in this study, the cooperative learning model and DMS were used together in the learning environment. When the codes of the findings of the qualitative data were investigated it was seen that DMS played a significant role. The reason why DMS was highlighted in the findings was so that the self-reported data obtained may enable students to focus only on this software.
Acknowledgments

This study was part of the first author’s doctoral dissertation, which was completed under the supervision of the second author. We would like to thank the students and teachers who participated in this study. In addition, we would like to thank the reviewers for their very helpful comments on our manuscript.

References


Herzig, A. H. (2002). Where have all the students gone? Participation of doctoral students in authentic mathematical activity as a necessary condition for persistence toward the PH. D. *Educational Studies in Mathematics, 50*(2), 177-212.


Şiştan, M., Lökçü, M., Oğuz, T., & Atak, Ö. (2013). *Secondary mathematics textbooks (Grade 11)*. Ankara, Turkey: Turkish Ministry of National Education.


Turkish Ministry of National Education [TMoNE], (2011). *Ortaöğretim matematik dersi (9, 10, 11 ve 12. sınıflar) öğretim programı* [Secondary mathematics curriculum (Grades 9-12)]. Ankara, Turkey: TMoNE.

Turkish Ministry of National Education [TMoNE], (2013). *Ortaöğretim matematik dersi (9, 10, 11 ve 12. sınıflar) öğretim programı* [Secondary mathematics curriculum (Grades 9-12)]. Ankara, Turkey: TMoNE.


Appendix 1

Worksheet – Example about sequences

Group name:
Group members:
Date: …………………
1) ………………………………………
2) ………………………………………
3) ………………………………………
4) ………………………………………
You can ask for advice from your teacher about the problems you have encountered in the worksheet while using the dynamic materials. Carry out your work with your group members. Please do not hesitate to write your opinions.

Let’s look at the M1 material.
Write the number of points in the table when the slider takes the values 1, 2, 3, …, 7

<table>
<thead>
<tr>
<th>n values</th>
<th>Number of points ((a_n))</th>
<th>What is the relationship between the (n) values and number of points?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Let \(f(1)=a_1, f(2)=a_2, f(3)=a_3, \ldots \) and \(f(n)=a_n\). Write the values of function \(f\) as \((a_n)=(a_1, a_2, a_3, \ldots, a_n)\).
Let \(f(n)=a_n\). Can we show any function \(f:N^+\to R\) as \((a_n)=(a_1, a_2, a_3, \ldots, a_n)\)? Explain.

Let’s look at M2 material.
You can get different visions by varying the sliders \(a\) and \(b\) in the graph-1 and graph-2 windows. What is the difference between the graph-1 and graph-2 windows? Explain.
Find whether the following functions \(N^+\to R\) are sequences or not.

a-) \(f(n)=2n+1\)

b-) \(f(n)=\frac{n}{n+1}\)

c-) \(f(n)=\frac{n}{n-2}\)

Which term of the sequence \((a_n)=\left(\frac{n-1}{n}\right)\) is equal to \(\frac{5}{6}\)?

Explain the difference between functions and sequences using graphs.
Screenshots of the materials M1 and M2, respectively that are used in worksheet-1 are given below.

**Figure 2. A screenshot of the material-M1**

**Figure 3. A screenshot of the material-M2**
Appendix 2

Worksheet – Example about quadratic functions

Group name:
Group members:
Date: …………………
1) …………………………………………………………
2) …………………………………………………………
3) …………………………………………………………
4) …………………………………………………………
You can ask for advice from your teacher about the problems you have encountered in the worksheet while using the dynamic materials. Carry out your work with your group members. Please do not hesitate to write your opinions.

Let’s look at the M3 material

Use slider \(a\) in order to examine the graph of \(f(x) = ax^2 + bx + c\). How does the slider impact the graph and the functions? Explain what are the relationships between the functions leading coefficient \((a)\) and the graphs of functions?

Use slider \(b\) in order to examine the graph of \(f(x) = ax^2 + bx + c\). How does the slider impact the graph and the functions? Explain what are the relationships between the values of \(b\) and the graphs of functions?

Use slider \(c\) in order to examine the graph of \(f(x) = ax^2 + bx + c\). How does the slider impact the graph and the functions? Explain what are the relationships between the values of \(c\) and the graphs of the functions?

Let’s look at the M4 material

Use slider \(a\) in order to examine changes that occur in the graph of \(f(x) = a(x - r)^2 + k\) in M4 material. Explain what are the relationships between the values of \(a\) and the graphs of functions?

Compare the peaks of various function graphs that are created from changing the leading coefficient (the values of \(a\)) of function \(f(x) = a(x - r)^2 + k\).

Ex: The functions \(f, g, h\) and \(k\) are in the form \(ax^2 + bx + c\) and in the each four graphics only “\(a\)” can take different values. Compare the values of \(a_f, a_g, a_h,\) and \(a_k\).

\[\text{Diagram showing graphs of } f, g, h, \text{ and } k.\]
Ex: The functions $f$, $g$, and $h$ are in the form $ax^2 + bx + c$ and in the each three graphics only “$c$” can take different values. Based on this compare the values of $c_f$, $c_g$, and $c_h$.

Ex: The graph of the parabola $y = a(x + 2)^2 + 1$ is given in below. Find the value of “$a$” in the graph.

Ex: The parabolas $f$, $g$, $h$, and $k$ are in the form $(f(x) = a(x - r)^2 + k)$ and in the each four graphics only “$a$” takes different values. Then:

a) Sort the descending the values of $a_f$, $a_g$, $a_h$, and $a_k$.

b) Compare peaks of the parabolas $f$, $g$, $h$, and $k$. 

87
Screenshots of the materials M3 and M4, respectively that are used in worksheet-2 are given below.

**Figure 4.** A screenshot of the material-M3

**Figure 5.** A screenshot of the material-M4
Investigating the Use of the Khan Academy and Mathematics Software with a Flipped Classroom Approach in Mathematics Teaching

Yılmaz Zengin
Department of Mathematics and Science Education, Dicle University, Diyarbakır, Turkey // yilmazzengin@outlook.com

(Submitted April 7, 2016; Revised August 26, 2016; Accepted October 8, 2016)

ABSTRACT
The purpose of this study was to determine the effect of the flipped classroom approach designed by using Khan Academy and free open source software on students’ academic achievement and to examine students’ views about this approach. The research was evaluated in the light of both qualitative and quantitative data. Twenty-eight students studying in the department of mathematics education in a state university in Turkey comprised the study group of the research which was conducted using a mixed methods research design. A double integral achievement test and an open-ended questionnaire about the flipped classroom approach were used as data collection tools. A Wilcoxon signed-rank test was used for the analysis of quantitative data and content analysis was used to analyse the qualitative data. According to the analysis of the research it was found that the flipped classroom approach designed with using both the Khan Academy and mathematics software increased student achievement in double integral. It was also found that this learning approach enhanced students’ understanding and provided visualization in mathematics teaching. Moreover, it was revealed that this approach promoted retention and made understanding much easier.

Keywords
Flipped classroom approach, Khan Academy, GeoGebra, Maxima, Double integral

Introduction
Information, communication and working styles have changed in the 21st century. This change has affected education and it has required computer and electronic technologies to be used in every field (Niess, 2005). Educators, particularly involving courses in mathematics that are difficult to understand (Freudenthal, 1983), have enabled students to better understand the concepts involved using technologies (Hoyles & Jones, 1998). The technologies have also given students opportunities to work on real life problems (Pierce & Stacey, 2011), and has also enabled them to identify different representations of concepts (Heid & Edwards, 2001).

It is regarded that using information communications technology (ICT) in very difficult mathematics courses is beneficial to students (Jones, 2000; Laborde, 1993; Marshall, Buteau, Jarvis, & Lavicza, 2012). Teachers use their knowledge of content, teaching and learning, and technology to promote experiences that develop students’ learning and creativity in computer-mediated environments (International Society for Technology in Education [ISTE], 2008). Moreover, teachers are expected to prepare their content by using a variety of software and to transfer them to the learning environment with the use of worksheets. In addition to teachers’ efforts in using these education technologies, it is important to consider how and with which approaches these technologies could be reflected in the classroom learning environment. It is considered that the flipped classroom, one of the blended learning models used widely, (Sahin, Cavlazoglu, & Zeytuncu, 2015) can enable teachers and students to structure the learning environment.

Essentially in a flipped classroom what is learned in class is learned at home, and homework done at home is now done in class (Bergmann & Sams, 2012). The traditional model of instruction is teacher-centred; the teacher gives lectures during the lesson and assigns students homework to do at home. The flipped classroom, or inverted classroom, reverses traditional education: the teacher delivers the content outside the classroom with videos prepared by him/her, and uses class time for active learning by having students collaborate and interact with each other (Mok, 2014). As a result of the flipped classroom, students find more opportunities to get engaged with more activities in class and to have discussions about the concepts involved. However, the teacher should very carefully plan activities, videos, presentations, or study notes to deliver content outside of the classroom.

There is also a concern that the flipped classroom can be regarded as one of the barriers between technology and teachers. However, Bergmann and Sams (2012) stress that the solution to overcoming the barriers in flipped classrooms is to employ, train, and support teachers. Moreover, although some critics fear that the Khan Academy’s importance can result in standardization and deprofessionalization, Bergmann and Sams (2012) and Andrea Smith point out that educational videos as important tools because teachers can develop content, share
resources, and promote practice (as cited in Tucker, 2012). The Khan Academy provides numerous activities, instructional videos, and a personalized learning dashboard that enable students to study at their own pace in and outside of the classroom. The Khan Academy guides students from nursery class to advanced mathematics by using the most developed and adaptive technologies. Moreover, the educator dashboard offers a summary of class performance. The Khan Academy founded by Salman Khan has grown into an 80-person organization that aims at providing a free world-class education for anyone, anywhere (Khan Academy, 2016). The Khan Academy is translated into different languages and offers content suitable to all levels in an entertaining environment by taking into consideration students’ knowledge gaps (Dijksman & Khan, 2011). It may be a suitable platform for students’ studies outside of the classroom in the flipped classroom approach. Although the duration of the videos in the Khan Academy is not very long, usually lasting between seven and 14 minutes, they provide opportunities for students to identify, explain, and practice different mathematics concepts using relevant software (Thompson, 2011).

In the flipped classroom approach, students benefit from the rich content of the Khan Academy at home before the actual class session, but it becomes more important that the kind of learning environment they will encounter in the classroom under the guidance of the teacher will enable them to enhance their understanding of mathematical concepts. If the preparations done by students outside of the classroom are not complemented with planned activities and with the teachers’ guidance in class, the positive effects of the flipped classroom may not emerge as expected. Thus, even if students understand the general framework of the concepts outside of the class, in order to examine the different representations and different situations of the concepts, the teacher is required to come to the class prepared. When the courses in the Khan Academy are examined, it is seen that some contents related to mathematics are enriched by using different software. Thus, students must be offered contents with relevant software. They must also be given opportunities to have different experiences with the software and to reinforce the concepts they have learned outside the class by means of discussions with their peers under the guidance of the teacher in the classroom environment. There are numerous open source software programs which promote the design of such learning environments in mathematics teaching. Maxima (Gaertne, 2005), GeoGebra (Hohenwarter & Fuchs, 2004), SageMath (Stein, 2006), Cadabra (Peeters, 2007), GNU Octave (Eaton, 1997) and Axiom (Paffman, 2007) are examples of such systems. Open source programs offer alternative contents to students and teachers for designing the learning environment.

The flipped classroom can provide higher order exploration opportunities for students if the activities in which students engage in the class are well-planned under the guidance of the teacher. Thus, benefiting from mathematics software with in-class activities can promote the efficiency of the flipped classroom strategy. In this study, Maxima and GeoGebra were actively used during the in-class implementations of the flipped classroom. Maxima and GeoGebra were preferred as open source mathematics software because they offer a variety of languages and are easy to use and set up. It is considered that students’ experiences about the flipped classroom approach can contribute to mathematics research and teaching. Few studies have been carried out with the flipped classroom approach in mathematics education (Love, Hodge, Grandgenett, & Swift, 2014; Moore, Gillett, & Steele, 2014; Sahin, Cavlazoglu, & Zeytuncu, 2015). These studies have provided limited explanations on how in-class and out of class activities of the flipped classroom approach were designed. Moreover, these studies resulted from out of school settings like in the Khan Academy and YouTube; however, using mathematics software which has a positive effect on learning in the classroom effectively has been mentioned in very few studies. In this study, the students who used the Khan Academy videos in out of class settings learned mathematics in informal cooperative activities by focusing on the double integral subject which they had difficulties in understanding. The flipped classroom approach supported with the efficient use of the Khan Academy materials in out of class settings and with mathematics software in in-class learning is important as it enables students to benefit from technology effectively while learning concepts with higher levels of difficulties. Moreover, the flipped classroom approach in mathematics education is usually used in middle and high school levels (Chen, Yang, & Hsiao, 2015; Fulton, 2012; Heo & Choi, 2014) but there is lack of research at the university level. The studies at university level are usually carried out involving students’ perceptions and attitudes and instructors’ experiences about the flipped classroom approach (Cilli-Turner, 2015; Love, Hodge, Grandgenett, & Swift, 2014; Ogden, 2015; Palmer, 2015). Due to the limited number of studies that have been conducted on students’ achievement, Ogden (2015) highlights that the flipped classroom approach needs to be examined in greater depth and its effects on students’ achievement should be investigated. Therefore, the effects of using the Khan Academy materials within the flipped classroom approach, by using a free and an open source mathematics software, were examined. In particular, the study involved the topic of the double integral, one of the topics which students have difficulties with and is considered to be one of the most important topics of calculus (Mathews, 1990). In this context, the purpose of the study was to determine the effect of the flipped classroom approach designed using the Khan Academy and mathematics software together on students’
achievement and to examine students’ views about this approach. With this aim, the research questions of the study are as follows:

- Research question 1 (RQ1): What is the effect of the flipped classroom approach designed by using the Khan Academy materials and mathematics software on student achievement in the double integral topic?
- Research question 2 (RQ2): What are students’ views on the use of the Khan Academy materials and mathematics software with the flipped classroom approach in mathematics courses?

Method

The study was carried out using a mixed methods design involving both quantitative and qualitative data. The design aimed at gathering both qualitative and quantitative data simultaneously or sequentially. The reason for gathering qualitative data is to supplement the primary quantitative data. Most examples in the literature reveal that qualitative data are embedded within the quantitative data (Creswell, 2012). In this study, the effect of the flipped classroom approach designed using the Khan Academy materials and mathematics software on students’ achievement was revealed by the quantitative data. The quantitative part of the research study was designed with a single-group using a pre-test and a post-test. Before the flipped classroom approach was implemented, a pre-test was administered. A post-test was then administered after implementation of the flipped classroom. Both the pre-test and post-test were the same tests (McMillan & Schumacher, 2010). Qualitative data were used to explain the students’ views on the flipped classroom approach.

Study group

Twenty-eight students studying in the department of mathematics education in a state university in Turkey comprised the study group of the research. The student consisted of 10 males and 18 females. Participants’ ages ranged from 22 to 26 years. The convenience sampling method was used in selecting the study group. In convenience sampling, subjects are selected because of their accessibility or convenience. Convenience samples are commonly used in both quantitative and qualitative studies; many researchers prefer this sampling technique because of its easy accessibility, efficiency, and is free from practical constraints (McMillan & Schumacher, 2010).

Data collection tools

The Double Integral Achievement Test (DIAT) and an open-ended questionnaire about the flipped classroom approach were used as data collection tools. The DIAT was developed taking into account the opinions of two mathematics educators and two mathematics experts. Moreover, the questions in the DIAT were taken from Thomas, Weir and Hass (2011), Balci (2009) and Balci (2010). Some of the questions in the DIAT are as follows:

- Is it possible to calculate the integral of a continuous function \( f(x,y) \) on a rectangular domain in the \( xy \) plane and obtain different results depending on the order of integration? Explain your answer with your reasons.
  For the following integral:
  \[
  \int_{0}^{2} \int_{0}^{2} (2x + 1)dydx
  \]

- Draw the integration domain;
- Write an equivalent integral by changing the order of integration.

The DIAT was evaluated using a holistic rubric. This rubric was developed by the researcher with help from the Turkish Ministry of National Education [TMoNe], (2013). The maximum score that could be obtained in this achievement test was 30. The qualitative data in the research were collected using an open-ended questionnaire. When the open-ended questionnaire was developed, two mathematics educators’ opinions were taken into account and a pilot test was carried out. After piloting, this questionnaire was finalized. One of the questions included in the open-ended questionnaire was: “What are the differences between your views before and after the implementation of the flipped classroom approach in the study of double integrals? Please explain.”

Implementation

The research study lasted eight weeks. In the first five weeks, students learned how to use GeoGebra and Maxima at a basic level. In the remaining three weeks, the topic of double integrals was taught using the flipped
classroom approach. In the first five weeks, implementations were carried out in 10 course hours (10 x 50 min.), with 2 hours in class per week and in the last three weeks. Instruction lasted for a total of nine course hours (9 x 50 min.), with three hours per week. The implementations were carried out in a total of 19 course hours (19 x 50 min.). Students watched the double integral videos from the Khan Academy outside of the class for three weeks before coming to class. Students studied in the computer lab using the worksheets prepared by the researcher in each lesson. Lessons in the computer lab are shown in Figure 1.

The researcher played the role of a guide when students used the mathematics software and worksheets. The researcher prepared three worksheets and four dynamic materials. The materials developed using mathematics software or by the students in the class were used with the worksheets. The worksheets were designed to provide opportunities for students to build and use materials. While developing the worksheets and materials, the websites http://tube.geogebra.org/ and http://maxima.sourceforge.net/ and Balcı (2009), Balcı (2010) were referred to. Hohenwarter and Hohenwarter (2012), Ku (2015), Thomas, Weir and Hass (2011) were also referred to. After the worksheets and materials were developed by the researcher, two mathematics educators who were experts in this field were asked to review the materials and the materials were revised based on feedback provided. Appendix 1 consists of some examples from the worksheets and related materials. A sample of the materials used in the lesson is given in Figure 2.

Here, using a slider, students were asked to change the n values. As the n value increased, they were asked to explain the relation between the volume of the solid obtained with the Riemann sum. An example of a material obtained with the increase of n-value is given in Figure 3.

During the lesson taught using the flipped classroom approach, informal cooperative activities were conducted in stages when students used mathematics software using the worksheets. Spontaneous group discussions and the think-pair-share technique were used in the informal discussions. The researcher asked students at various times during the lesson what mathematics concepts mean, why something works and how a mathematics problem may be solved in a spontaneous group discussion. This group discussion was used during the first week. In the second and third weeks, the researcher used the think-pair-share strategy. The students paired with another student within their teams. The researcher posed questions using the mathematics software and worksheets to the class. The students were required to think of an answer on their own, then to pair with their partners to concur on a
solution. In the end, the researcher asked students to share their answers or thoughts with the class (Slavin, 1995). Thus, students were expected to come to the lesson prepared with Khan Academy materials and under the guidance of the researcher students had more opportunity to focus on and discuss the topic of double integrals.

**Figure 3.** With the increase of n-value, the Riemann sum approximations approach the total volume of the solid

### Data analysis

The data of the research were obtained from the DIAT. Normality analysis of the quantitative data obtained from the DIAT was performed. The Shapiro–Wilk test was performed and Q-Q plot, box and whisker plot, kurtosis and skewness coefficients were examined (Field, 2009). The results of the Shapiro–Wilk test were as follows ($p_{\text{pretest-DIAT}} < .05; p_{\text{posttest-DIAT}} < .05$). Analyses of the test results, as well as the Q-Q plot, box and whisker plot, kurtosis and skewness coefficients revealed that both the pre-test and the post-test did not show normal distribution. For this reason, the Wilcoxon signed-rank test was used to examine the differences between the pre-test and the post-test. The SPSS 18.0 program was used for the data analysis. Regarding the significance level, the value of .05, which is most frequently used in educational studies, was considered. For the calculation of the effect size, the $r = z / \sqrt{N}$ equation was used (Field, 2009), while the following criteria for the values of $r$ [$r = .10$ (small effect), $r = .30$ (medium effect) and $r = .50$ (large effect)] were considered for the effect size (Cohen, 1992).

Content analysis was used for the analysis of the qualitative data. Categories were created by coding student responses to each question in the open-ended questionnaire. These were presented in tables with the frequencies and percentages. Moreover, regarding the categories created, the sample quotations taken from the students were coded as S1, S2, S3 ... S28.

### Findings

In the research, the DIAT was used as a pre-test before the experimental treatment and as a post-test after the experimental treatment to determine the effect of the flipped classroom approach designed with using Khan Academy materials and mathematics software for the implementation of the topic of double integrals on students’ academic achievement. The Wilcoxon signed-rank test was used to examine whether or not there was a significant difference between the scores students achieved in the pre-test and post-test. Students’ test results are given in Table 1.

The analysis results presented in Table 1 demonstrate that the scores students achieved from the double integral achievement test after the implementation of the flipped classroom approach ($Mdn = 23$) were at a significantly higher level than before the implementation ($Mdn = 3$) scores of the test, $z = -4.21, p < .05, r = .62$. Moreover, when the mean scores of the pre-test and post-test were examined, students’ pre-test mean score was 1.69. After the intervention students’ post-test mean score was 21.82. Based on these results, it can be stated that the flipped classroom approach designed using the Khan Academy materials and the mathematics software was an effective approach to increase students’ achievement.
Table 1. Wilcoxon-signed rank test results of the students’ pre-test and post-test scores

<table>
<thead>
<tr>
<th>Post-test pre-test</th>
<th>N</th>
<th>Mean rank</th>
<th>Sum of ranks</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative rank</td>
<td>0</td>
<td>.00</td>
<td>.00</td>
<td>-4.212*</td>
<td>.00</td>
</tr>
<tr>
<td>Positive rank</td>
<td>23</td>
<td>12.00</td>
<td>276.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ties</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *Based on negative ranks.

As a result of implementation of the flipped classroom approach, three categories called “advantages,” “students’ suggestions,” and “disadvantages” were created from the analysis of the data obtained from the open-ended questionnaire responded to by the 28 volunteer students. Table 2 presents the codes and the categories belonging to these codes as well as the frequencies (f) of students involved.

Table 2. Evaluation of the approach

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>f</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
<td>Enhancing understanding</td>
<td>22</td>
<td>78.57</td>
</tr>
<tr>
<td></td>
<td>Visualization</td>
<td>17</td>
<td>60.71</td>
</tr>
<tr>
<td></td>
<td>Promoting retention</td>
<td>17</td>
<td>60.71</td>
</tr>
<tr>
<td></td>
<td>Making understanding much easier</td>
<td>15</td>
<td>53.57</td>
</tr>
<tr>
<td></td>
<td>Coming to the class prepared</td>
<td>13</td>
<td>46.42</td>
</tr>
<tr>
<td></td>
<td>Concretization</td>
<td>8</td>
<td>28.57</td>
</tr>
<tr>
<td></td>
<td>Making difficult concepts easy</td>
<td>8</td>
<td>28.57</td>
</tr>
<tr>
<td></td>
<td>Liking the topic</td>
<td>7</td>
<td>25.00</td>
</tr>
<tr>
<td></td>
<td>Moving away from memorization</td>
<td>6</td>
<td>21.42</td>
</tr>
<tr>
<td></td>
<td>Understanding faster</td>
<td>6</td>
<td>21.42</td>
</tr>
<tr>
<td></td>
<td>Promoting productive learning</td>
<td>4</td>
<td>14.28</td>
</tr>
<tr>
<td></td>
<td>Increasing motivation</td>
<td>4</td>
<td>14.28</td>
</tr>
<tr>
<td>Students’ suggestions</td>
<td>Using Khan Academy and mathematics software together</td>
<td>23</td>
<td>82.14</td>
</tr>
<tr>
<td></td>
<td>Must be implemented in all subjects</td>
<td>7</td>
<td>25.00</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>The inadequacy of using computers</td>
<td>5</td>
<td>17.85</td>
</tr>
<tr>
<td></td>
<td>Difficulties in using mathematics software</td>
<td>2</td>
<td>7.14</td>
</tr>
</tbody>
</table>

When Table 2 is analysed in terms of the “advantages,” it emerges that the flipped classroom approach designed using the Khan Academy materials and mathematics software together enabled students to understand the concepts better. Similarly, it was emphasized by the students that the approach promoted visualization and retention of the knowledge learned and it fostered learning of concepts more easily. In addition to these, Table 2 suggests that the approach helped students to come to the lesson prepared, it concretized the subject and also students loved it while at the same time it moved students away from memorization.

Most of the students stated that the flipped classroom approach enabled them to understand the subject better. The views of S7 can be given as an example:

“Regarding the preparedness level, we come to the class in a much better way and we watch the subjects which we are going to study via computer on videos and then we study them in the class. This has a positive effect on my learning. I noticed that I comprehended much better.”

The reason why students understood the concepts much better with the flipped classroom approach designed using the Khan Academy materials and mathematics software is that this approach provides more visuals in mathematics courses and students are prepared before coming to the class. The views of S2 and S25 supporting this opinion can be given as examples:

“When subjects are supported with visuals, they became more understandable. I could construct the examples much more easily with my logic...” (S2)

“I believe that this model has been very helpful for us to come to the lessons prepared. Because it enhances the level of our prior knowledge, it enables us to learn more clearly and permanently...” (S25)

A total of 60% of the students stated that the approach promoted visualization and retention of knowledge. The following views of two students can be given as examples.

“As it is supported with visuals, it is remembered easily and it moves us away from memorization.” (S2)

“As we see visuals, they are remembered easily and I do not forget what I see. I am so lucky about this issue. It enables me to remember easily. I understood double integral very well and I will never forget it.” (S10)
When the views of S2 and S10 are examined, it is revealed that in this approach, the visuals provided before the lesson and during the lesson promoted retention of the concepts learned. Two students’ views supporting this can be given as follows:

“It was proved once again that the visuals enhanced retention. I could not solve integral before but now I can solve double integral… Long live flipped mathematics classroom.” (S9)

“Examining mathematics and viewing it in detail (with the support of 3D) on computer environment increase.” (S11)

As many as 53% of the students stated that they understood the concepts much more easily with this approach. The following views of two students can be given as examples:

“Before the implementation, I was so scared. After the implementation, I understood that the subject was easy…I will do it with inner peace.” (S14)

“Thanks to this model, we understood that those subjects which are difficult to comprehend like double integral would be apprehended more easily.” (S22)

Only 46% of the students stated that they came to the class prepared because of the approach. S28 shared his opinion as follows:

“This model enables students to be active in the learning process. Thus, more permanent learning is actualized. In addition, coming to the class prepared for the subject which we will learn in the class increases more interest in the course. It promotes motivation.”

Considering the views of S28, it may be concluded that students were of the opinion that their motivation towards the course will increase when they come prepared to the lesson.

Only 28% of the students stated that the approach concretized the subject and 25% of them mentioned that the approach made them like the subject. S5 and S9 shared their views on the topic:

“I think that it has benefits for concretizing abstract concepts.”

“Coming to the class prepared helps me to eliminate the prejudices against the course. Thus, I liked integrals. Integrals .... Makes me happy.”

When students’ views were examined, it was revealed that the concepts before the course became more concrete with this approach and this caused students to learn the subject more easily. The views of S13and S19 supporting this opinion can be given as examples:

“We have to come to the lesson prepared with this model. This enables us to understand much more easily and quickly.” (S3)

“Usually students have difficulty with forming concepts. Because flipped mathematics classroom and software turn the abstract concepts into concrete ones, it is much easier to understand them...” (S19)

Only 28% of the students stated that it was possible to learn difficult mathematics concepts simply with the flipped approach designed using Khan Academy materials and mathematics software. The views of S16 can be given as an example:

“It teaches how to learn as it presents the complex subjects in a simple way. As I told, complexity is reduced to simplicity...”

Only 21% of the students stated that the approach moved them away from memorization. S3 shared her opinion that:

“I learned the double integral rationally because of the flipped mathematics classroom. It is so clear that it is against rote learning and lecturing.”

Students stated that the visualization provided with this approach promoted meaningful learning rather than rote learning. A students’ view can be given as an example:
“Khan Academy visualises learning and concretizes abstract concepts... The abstract and difficult concepts to conceptualize in our mind can be seen easily and understood better with mathematics software.... Thus, flipped model is more effective in meaningful learning due to understanding and visualizing than rote learning.” (S26)

When Table 2 was examined regarding “suggestions,” most of the students stated that in the flipped classroom approach, the Khan Academy materials and mathematics software should be used together. Moreover, the students stated that they wanted this approach to be used in all subjects. The following views of two students belonging to this category can be given as examples.

“Khan Academy teaches learning because it teaches the complex subjects in a simple way... The model and Khan Academy are ideal to teach difficult subjects easily. As I said, complexity is reduced to simplicity. There is only one thing missing here. The lack of visuals (graphs, diagram, etc.) are completed by GeoGebra and Maxima. All three (GeoGebra, Maxima and Khan Academy) go well together.” (S4)

“It is a very good model. In my opinion, it must be used with all subjects.” (S2)

When the “disadvantages” were examined in Table 2, five students stated that they had difficulty due to lack of information about how to use a computer while two students stated that they had difficulty with using mathematics software. Two students’ views can be given as examples of this category:

“I had difficulty because of lack of information about how to use a computer.” (S22)

“I had difficulty with generating graphics with software.” (S26)

Discussion

Within the context of this study, students successfully used the Khan Academy materials related to the use of double integrals outside the classroom. They learned about the double integral in class through informal cooperative activities using materials and worksheets which the researcher had designed and developed from the literature using GeoGebra and Maxima. Quantitative and qualitative data were collected during the implementation of the approach in the research. The analysis of qualitative data demonstrated that students understood the concepts much better as a result of using the flipped classroom approach that was designed using the Khan Academy materials and mathematics software together. In addition, it was emphasized by the students that the approach helped them to visualize the concepts, promoted retention of knowledge, and fostered easier learning of the concepts. Moreover, it was revealed that the approach helped students to come to the lesson prepared, it concretized the subject and made students like the subject. These findings are similar to the results of the study carried out by Sahin, Cavlazoglu and Zeytuncu (2015).

Another finding from the qualitative data was that the approach might move students away from memorization. Considering the views of S3, S25, and S28, it was determined that coming to the class ready with this approach enhanced students’ understanding of the subject better and much easier and increased their motivation. Moreover, it was determined that with this approach, using worksheets in class with the Khan Academy materials and the mathematics software together made lessons more visual and concrete for students. From this point of view, it can be stated that students understood mathematics much better and more easily since the well-prepared students discussed the concepts more concretely and visually. It was found based on the views of S9 and S11 that the visuals obtained with this approach promoted retention of knowledge. Teaching concepts in a more visual environment might move students away from memorizing and learning can be actualized in a more conceptual learning environment. Especially the view stated by S26 who emphasized that by using this approach, mathematics was learned in a meaningful way rather than by memorizing, supports this opinion. When the qualitative data obtained were examined, it was found that well-prepared students studied mathematics in a more concrete and visual environment and thus, understood concepts much better and more easily. The students’ views revealed that visualization offered opportunities for students to learn the concepts more permanently and also to move away from memorizing; it provided a meaningful learning environment. It can be stated that these positive contributions gained with this learning approach increased students’ achievement in the learning of the double integral. The quantitative data obtained revealed that the flipped classroom approach designed using the Khan Academy materials and mathematics software together promoted student achievement. The results obtained in this research correspond to other studies that have been conducted on the flipped classroom approach in mathematics teaching (Fulton, 2012; Love, Hodge, Grandgenett, & Swift; 2014; Sahin, Cavlazoglu, & Zeytuncu, 2015; Van Sickle, 2015). The evaluation carried out based on both quantitative and qualitative data obtained in the research study are given in Figure 4.
The flipped classroom approach designed using the Khan Academy materials and mathematics software provided students with the opportunity to study mathematics in a visual and concrete learning environment. Students who came to the class were prepared to take the opportunity to study mathematics in a concrete environment using static concepts in a dynamic and visual environment with the support of the Khan Academy materials and mathematics software. Thus, the difficult mathematics topics are taught in a simpler way. S16 emphasized that she learned about the double integral using this method while she was previously having great difficulty in understanding the definite integral concept. The flipped classroom approach supported with the Khan Academy materials and mathematics software enabled students to understand mathematical concepts much better and more easily and thus students’ achievement increased. Although this approach made positive contributions, some students had difficulties due to their failure to use a computer and software. Therefore, it is considered that the implementation of this approach in the classroom setting becomes a disadvantage for students who do not know how to use a computer.

Conclusion and suggestions

The study examined the effect of the flipped classroom approach designed using the Khan Academy materials and mathematics software on student achievement of the topic of double integrals and elicited students’ views about this approach. With regard to RQ1 (What is the effect of the flipped classroom approach designed by using the Khan Academy materials and mathematics software on student achievement in the double integral topic?), the qualitative and quantitative data obtained demonstrate that this approach is an effective model in increasing student achievement. In addition, when not only the differences between pre-test and post-test scores but also the mean ranks were compared, it can be stated that this approach increased student achievement. Moreover, based on students’ views with respect to RQ2 (What are students’ views on the use of the Khan Academy materials and mathematics software with the flipped classroom approach in mathematics courses?), it was found that this approach enhanced students’ understanding of the mathematical concepts, it made the course more visual, and it promoted retention. Besides the three main contributions of this approach, it was revealed that this approach made understanding much easier. Therefore, the contributions of the approach may promote the achievement of students.

In this research, the flipped classroom approach was supported with Khan Academy materials outside of the classroom and with free open source software like GeoGebra and Maxima that were used in the class. It was found that this approach supported with using the Khan Academy materials and free open source software together can be an effective model for a student anywhere in the world to increase his/her achievement in a course with a high level of difficulty like mathematics. Teachers who want to implement the flipped classroom approach in their lessons can benefit from the Khan Academy materials instead of preparing the contents themselves. There is a lot of free open source software offered for use in class activities. GeoGebra and Maxima used within the context of this study offered a dynamic learning environment to students thus providing opportunities for visualization and concretization. The teachers who are going to use the flipped classroom approach can benefit from the Khan Academy materials and free open source software together.
approach can enrich their instruction with such software. So, if the flipped classroom approach is not supported with such software programs and worksheets, the strengths of the approach and its contribution to students’ understanding may not emerge. It can be much easier for students who study the course outside of the class with platforms like the Khan Academy to enrich their conceptual knowledge using mathematics software in class. Therefore, it is suggested that the flipped classroom approach should be used supported with such software programs.

Limitations

The research has two fundamental limitations. One of them is related to the quantitative side of the research. The quantitative part of the study was designed with a single-group. Because there was only one group, comparison was made only within the group. If a randomized pretest-posttest control group design or a non-equivalent pretest-posttest control group had been used, the data of the experimental group would have been compared with the control group data. Another limitation of the study is that the study group was determined by the convenience sample method. As the researcher does not have the opportunity to work with a large number and diversity of participants, this limitation has emerged. Furthermore, the self-reported data presented here may cause students to focus only on limited aspects of the approach. As seen in the results, students mostly focused on positive effects of the approach rather than negative effects. While this enriched the results of the study by highlighting the advantages of the study, it also limited identifying disadvantages of the approach.

References


Appendix 1

An Example of Worksheets Contents

You can ask for advice from teacher about the content and the software. Carry out your work with your pair.

- Using GeoGebra, explain Fubini’s Theorem. Use input, graphics, and GeoGebra tools. In addition, you can use 3D graphics.
- Open material M2. Interpret how to obtain the volume in double integrals with the help of material M2.
- State the integration region of the following integral with the help of GeoGebra where R is bounded triangle with x-axis, y=x and x=2 lines

\[ \iint_R f(x, y) \, dA \]

- On the region R, find integration bounds for \( \iint_R dA \):
  - by using vertical lines
  - by using horizontal lines

![Figure 5. A screenshot belonging to the worksheet](image)

- Find the volume of the region which is bounded above by the paraboloid \( z = x^2 + y^2 \) and bounded below by the triangle formed from the lines \( y=x \), \( x=0 \) and, \( x+y=1 \). Verify the obtained result with maxima. Draw the surface with maxima

- For the following integral

\[ \int_0^3 \int_0^{3/2} 8x \, dy \, dx \]

  - sketch integration region
  - write an integral which is equivalent to integral whose integration order is changed
  - find the value of integral with maxima

- Define a surface with the help of maxima and find the volume below the surface in the integration bounds.
How Does Mozart’s Music Affect Children’s Reading? The Evidence from Learning Anxiety and Reading Rates with e-Books

Yen-Ning Su¹, Chih-Chien Kao¹, Chia-Cheng Hsu², Lu-Chun Pan³, Shu-Chen Cheng⁴ and Yueh-Min Huang¹

¹Department of Engineering Science, National Cheng Kung University, Taiwan, ROC // ²Department of Visual Communication Design, Tainan University of Technology, Taiwan, ROC // ³National Feng Hsin Senior High School, Taiwan, ROC // ⁴Southern Taiwan University of Science and Technology, Taiwan, ROC // yennen@mail.tn.edu.tw // njiingm@gmail.com // chia.cheng.hsu.tw@gmail.com // darkk6@gmail.com // kittyc@stust.edu.tw // huang@mail.ncku.edu.tw

*Corresponding author

(Submitted October 2, 2015; Revised June 29, 2016; Accepted September 27, 2016)

ABSTRACT

Some music has been proved effectively to mitigate anxiety, beneficial to reading. However, little was known on its influence of reading behaviors and outcomes. Thanks to the advance of e-book technology, it becomes possible to track reading rate and outcomes in a real-time manner by the underlying mobile devices. This study intends to examine the effects of the Mozart piece K.448 (sonata for two pianos in D major) on the learning anxiety, reading rates and reading comprehension of students for reading e-books. The quasi-experimental design was applied, and 62 senior grade elementary school students participated in this experiment. The results showed that when compared with a silent task (reading without the music), this piece of Mozart’s music had positive effect in reducing learning anxiety, and improving the students’ reading rates, reading comprehension and direct process performance. However, we found that the music had a negative effect on the students’ attention for their interpretation process. We inferred that this Mozart’s music might take part of attention resource. This was supported by the increase of students’ extraneous load while listening to Mozart music, which led to the insufficient concentration in the interpretation and the evaluation of content. Overall, the findings reveal that the use of certain music, such as Mozart K.488, can be a potential tool to enhance reading outcomes when using e-books.

Keywords
Mozart music, Learning anxiety, Reading rate, Reading comprehension, E-books

Introduction

Reading is the foundation of lifelong learning, and is an important indicator of national competitiveness (Lin, Liu, Lin, & Li, 2008). In recent years, the Taiwanese government has been committed to promoting the implementation of primary and secondary school reading programs (MOE, 2007), with the aims of improving primary and secondary students reading abilities, and creating a user-friendly environment to attract students’ interest in reading (MOE, 2007), which is essential for enhancing reading comprehension. Some studies have explored ways to improve reading abilities through stimulating students’ interest in reading, such as the use of reading strategies (Khowaja, & Salim, 2013), curricula and instruction (Schüenemann, Spörer, & Brunstein, 2013; Huang, & Chiu, 2015a; Huang, & Chiu, 2015b), and the use of technology (Huang, & Liang, 2015; Ponce, López & Mayer, 2012; Liang & Huang, 2014; Huang, Huang & Wu, 2014). The development of e-books, which are usually equipped with multimedia functions, has attracted the attention of many educators (Huang & Liang, 2015). The Ministry of Education of Taiwan has carried out a series of pilot plans for e-books (MOE, 2009, 2010), in hope of enhancing the effectiveness of teaching and learning through related technologies. Many scholars have shown the positive effects that can be achieved by introducing e-books to teaching or learning practice (Korat & Shamir, 2008). It is thus expected that instructors should be able to take advantage of e-books to build a better or more innovative reading environment.

Stages of reading development and anxiety

Chall (1983) stated that the development of children’s reading may be segmented into five stages, with 9-14 year-old students in the “Reading for Learning New Information” stage. Students in this stage usually have a single viewpoint when reading, and obtain new information from a reading process. Gunning (1996) claimed out when high-grade elementary school are in this stage the purpose is not only to learn how to read, but also for them to expand their knowledge based on the reading methods used. Therefore, textbooks used by higher-grades are longer and more difficult than those used by lower-grades. Wu and Lai (2007) noted that anxiety affects
students’ learning processes, and excessive anxiety would interfere with their cognitive processes (Wood, 2006), leading to relatively poor learning outcomes. Since they are dealing with more difficult learning material, higher grade students tend to encounter more challenges that are different from those seen in their past reading experiences, and thus may produce a considerable degree of anxiety. However, some studies have claimed that an appropriate level of anxiety can promote learning, with only excessive anxiety having negative effects (Chang, 1999; Vickers & Williams, 2007). Although some studies have shown that e-book can have a positive impact on reading abilities when integrated into reading instruction (Korat & Shamir, 2008). Additionally, Lam et al. (2009) emphasized that the phenomenon of eBooks is quite different from the traditional books in use today. Wu et al. (2011) also argued that the use of such technology might compromise learning outcomes, since students have to learn how to operate in different learning contexts, regardless of an individual differences with regard to technology acceptance or adaptive ability. Therefore, how to effectively reduce the anxiety of higher grade elementary students, and further promote the development of their reading abilities by the use of e-books, are two issues worth exploring.

Use music to reduce anxiety

As anxiety can harm an individual’s performance (Vickers & Williams, 2007), some studies have explored the effects of using music to mitigate anxiety (Shen, Wang & Shen, 2009). For example, Qin et al. (2014) demonstrated that listening music as a way can compensate negative emotion of students. Dolean’s (2015) study explored the effect of music on students’ anxiety during the regular language classes, and found that music could reduce students’ anxiety. In order to assess the effects of music on examination anxiety, Lai et al. (2008) conducted a series of experiments with thirty-eight students randomly assigned to a music or a silent group. The students in the music group listened to a slow piece of music and then received a test, while the students in the silent group took a test without listening to any music beforehand. The results showed that the anxiety of the experimental group was lower than that of the control group, implying that slow music helps reduce anxiety.

Dusseville et al. (2012) investigated the influence of music on 249 undergraduate students. The experimental group students listened to classical music during a lecture, while the control group listened to the lecture without music. The results indicated that the performance of the experimental group was significantly higher than that of the control group, indicating that classical music has a positive influence on learning performance. de Groot (2006) also found that more learning occurred in the music condition (classical music) than in the silent condition when students were learning language vocabulary. Hall (1952) pointed out that reading comprehension among eighth and ninth graders improved in the presence of background music. Rauscher et al. (1993) explored the impact of using Mozart’s sonata for two pianos, K.488, on the performance of undergraduate students. The findings showed that this music could help the students’ spatial reasoning abilities, with average score of the experimental students being 10 points higher than that of the control group. A phenomenon called the Mozart effect was thus proposed, which lead to a number of subsequent studies. For instance, Campbell (1997) also used the same Mozart’s piece to help patients reduce anxiety. Jaušovec et al. (2006) found that this music could also help to enhance students’ learning. These earlier studies suggest that Mozart’s sonata for two pianos, K.488, has a positive impact on reducing anxiety and promoting individual learning. In addition, Shih et al. (2012) pointed out that, compared to adagio or classical music, music that contains lyrics may have a negative impact on learning concentration and efficiency. Pring and Walker (1994) noted that music with lyrics interferes with other brain processes. Xing et al. (2016) found that rhythm of Mozart’s piano sonata K.448 is the crucial factor of Mozart effect. Besides, Xing et al. (2016) also suggested that different music may have quite different to opposite effects. Therefore, in order to mitigate the anxiety of students and enhance their performance, teachers should consider using appropriate background music, such as s Mozart’s sonata for two pianos, K.488.

However, music can also be a negative factor to students’ learning, learning environment and learning anxiety (Dolean, 2015) when the instructor integrates music into regular learning activities. Sweller (2010) claimed that cognitive load theory could be distinguished between two sources of cognitive load, namely intrinsic and extraneous load. Reedy (2015) emphasized that the intrinsic load of a learning environment or task is concerned with its inherent difficulty for a learner. Moreover, either poorly designed learning experiences or learning environment will increase extraneous load (Reedy, 2015). Based on cognitive load theory, Schellenberg (2005) claimed that background music may affect human’s cognitive load and task. Thus, the effect of background music on cognitive load when students are reading e-books is an issue worth exploring.

Based on these earlier works, this study examined the three following research questions. (1) Can Mozart reduce the learning anxiety of students when they are reading e-books? (2) Does Mozart enhance the reading comprehension of students when they are reading e-books? (3) Does Mozart affect the cognitive load of students when they are reading e-books?
Anxiety, task efficiency and reading rate

Some studies have shown that high anxiety is associated with lower task efficiency (Tanaka, Takehara & Yamauchi, 2006). Byrne and Eysenck (1995) also found that the task efficiency of subjects with high anxiety was lower than that of the low-anxiety subjects. Seliger (1972) found that students with low levels of anxiety had faster reading rates. Gifford et al. (1966) claimed that anxiety is also seen in the behaviors of learners, with the reading rates being negatively related to the level of anxiety. For this reason, this study adopted Mozart’s sonata for two pianos, hereafter K. 488, in the learning activities to investigate its impact on the students’ anxiety, with the students’ reading rates used as the focal performance indicator. The traditional approach in such studies is to use cameras to record learning behaviors and reading rates. However, this method is time-consuming when tracking the learning histories of many students. In contrast, Huang and Liang (2015) used the touch screens of tablet computers to record the e-book reading rates of students, and the results showed that this was able to produce accurate results. This approach can only track reading rates in real-time, and enable instructors to easily see the indications of the learning anxiety of individual learners. The fourth research question examined in this work is thus: Is there a relationship between learning anxiety and reading rates when students are reading e-books without music?

Goals and hypotheses of the current investigation

As noted above, previous studies have found that Mozart’s K.488 can reduce anxiety and improve learning. This study thus explored the impact of this piece of music on the learning anxiety, cognitive load, reading rates, and reading comprehension of learners using e-books, based on the following hypotheses:

- **Hypothesis 1.** Learning anxiety will be reduced when students are reading e-books with Mozart’s K.488 as background music.
- **Hypothesis 2.** The intrinsic load measures are not significant between silent task and Mozart music task when students are reading e-books.
- **Hypothesis 3.** Extraneous load will be increased when students are reading e-books with Mozart’s K.488 as background music.
- **Hypothesis 4.** Reading comprehension will be enhanced when students are reading e-books with Mozart’s K.488 as background music.
- **Hypothesis 5.** Learning anxiety and reading rate are negatively correlated with each other when students are reading e-books without background music.

Method

Participants

At the beginning of the experiment, 66 senior-grade students (37 males and 29 females) in an elementary school in southern Taiwan were recruited to participate in the learning activity. By excluding those who did not finish the whole process, 62 sets of valid experimental data were collected for analysis. The details of the participants are shown in Table 1.

<table>
<thead>
<tr>
<th>Participants</th>
<th>All participants</th>
<th>All participants without invalid data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group I</td>
<td>Group II</td>
</tr>
<tr>
<td>Male</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>29</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

Experimental environment and tools

The reading materials were three articles of Chinese expository prose produced by the Taiwanese government for use in elementary schools (MOE, 2011), the titles of which were “Pigeon,” “The Glacier” and “Rock Climbing.” The first article (“Pigeon”) was used to conduct the pre-test. The second article (“The Glacier”) contains 1,206 words, and the average sentence contains 10.95 words, with 99.42% of these from the list of 5,021 common Chinese words for elementary school students (MOE, 2000). The third article (“Rock Climbing”) contains 1,199 words, with an average of 11.1 words per sentence, and 99.58% of these are from the same set of common words.
Both texts (“The Glacier” and “Rock Climbing”) were slightly above the students’ reading ability, in order to ensure that they were appropriate for use in this context with regard to the students’ word recognition abilities, thus preventing a ceiling effect. Furthermore, in order to control for the influence of extraneous variables, and record the students’ reading rates accurately while reading, all reading materials were text-only (excluding the cover).

We used the Android version of Office Suite installed on tablet computers as the reading software. Previous studies indicated that elementary school students often could not easily focus on the text, and some teachers adopt a pointing strategy to require the students to use a finger to follow what they are reading (Chen & Chang, 2011; Huang et al., 2014). Thus, a program was developed to follow the movement of each student’s fingers on the touch screen in order to measure the reading rate, this method also echoing the students’ previous reading experiences. To track the reading rates, all students were asked to use a finger to follow the text on the touch screen when they are reading, and the system then recorded the words read per minute.

In order to ensure that students of group I and group II have the same reading ability before learning, this study used a reading comprehension test (The first article: “Pigeon”) as pre-test to compare reading comprehension of the students of both groups and see whether their reading abilities are consistent. The comprehension test for “Pigeon” has 14 questions with a total of 21 points. The result of the pre-test shows that $t = 1.176$, $p > .05$, and the pre-test results of both direct and interpretation process shows that $t = 1.341$, $p > .05$; $t = 0.523$, $p > .05$, respectively. With no statistically significant difference, which indicates that group I and group II have the same reading ability before learning; therefore their reading abilities can be regarded as consistent.

In learning task, we used a reading comprehension test for each article (“The Glacier” and “Rock Climbing”) to examine students’ reading comprehension when students finished their learning tasks. The comprehension test for “The Glacier” has 13 questions with a total of 18 points, including seven multiple-choice questions (7 points) and six short-answer questions (11 points). The comprehension test for “Rock Climbing” has 14 questions for a total of 18 points, with seven multiple-choice questions (7 points) and seven short-answer questions (7 points). The difficulties of the questions for both tests range from 0.6 to 0.7, categorized as easy to mid-level difficulty. In other words, both comprehension tests were suitable for to assess the students’ reading comprehension in this study. In order to ensure scoring validity, we invited two experts, who are teachers in an elementary school, to serve as raters and assess the short-answer questions according to the scoring guides for each test. After the two raters examined the tests, we examined the inter-rater reliability using Pearson’s product-moment correlation coefficient. The inter-rater reliability ranged from 0.86 to 0.95 ($p < .01$), thus indicating good reliability (Chiu, 2009).

We revised learning anxiety scales from two previous studies (Venkatesh, 2000; He, Chang & Liu, 2010) to measure the participants’ learning anxiety. The scale consisted of eight questions that the students answered when they had finished each learning task. The responses to all questions were on a four-point Likert-scale, ranging from 4 for strongly agree to 1 for strongly disagree, with higher scores indicating less learning anxiety. The internal consistency and reliability were tested by the Cronbach’s alpha coefficient, and the result for the sample as a whole was .836, indicating that the scale was acceptable, with good internal consistency and reliability.

This study used the scale of cognitive load designed by Ouyang et al. (2010) to analyze the cognitive load of all students. The scale consisted of 5 items of questions that were used to access the score of students’ cognitive load when they finished each learning task. Responses to all questions were on a four-point Likert-scale, from 4 for strongly difficult to 1 for strongly easy, thus the score range of cognitive load was from 5 to 20 with higher scores representing higher cognitive load.

**Experimental design and procedure**

Learning anxiety and cognitive load were the situational variables of the subjects taking part in this study, and the learning anxiety and cognitive load of individual participants are correlated could not be determined prior to experimentation. With this limitation, this study used the quasi-experimental design. With the same number of subjects receiving each experimental manipulation to explore the impact of Mozart’s K.488 on learning anxiety and reading comprehension. However, the use of the same group of subjects in the experimental activities may cause some bias in the results, due to practice and aging effects, so we adopted the following strategies to reduce the impact of correlation effects: using texts and tests with the same difficulty to avoid the practice effect, and shortening the experimental time interval to one week to reduce the aging effect. The learning task with the Mozart background music was is referred to as the “music task,” and that without the music as the “silent task.”
In addition, we used two groups of participants with both groups following the same experimental procedure. In order to enhance the internal validity and reduce the impact of the differences of reading materials, each group was assigned to read the same articles but in different orders. In the first silent reading experiment, while group I read “The Glacier,” group II read “Rock Climbing.” One week later, in the music task, group I read “Rock Climbing” while group II read “The Glacier.” The experimental design is shown in Table 2.

Table 2. Experimental design

<table>
<thead>
<tr>
<th>Task</th>
<th>Test</th>
<th>Time interval</th>
<th>Task</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>T₁</td>
<td>O₁</td>
<td>X</td>
<td>T₂</td>
<td>O₂</td>
</tr>
</tbody>
</table>

Note. X: The interval time is a week; T₁: All participants read the material without Mozart; T₂: All participants read material with Mozart; O₁, O₂: Test (learning anxiety, cognitive load, and reading comprehension).

In order to explore the impact of Mozart’s K.488 on the students’ learning anxiety, cognitive load and reading comprehension, a series of experimental activities were used, and the details of these are shown in Figure 1.

Figure 1. Experimental procedure

Based on the experimental design, the experimental procedure was divided into two main phases: (1) Reading e-books without Mozart, and (2) Reading e-books with Mozart. Each phase contained reading and test tasks. At the beginning of the experiment the students were given instructions with regard to the assigned tasks to avoid the Hawthorne effect.

In the first phase, the silent task was 30 minutes of reading, in which group I was assigned to read “The Glacier,” and group II was assigned to read the “Rock Climbing.” Before the task, students were asked to read the article carefully and pay attention to the text so that they could clearly understand the main points. Upon completion of the reading task, the students were asked to complete the learning anxiety scale and a 20-minute reading comprehension test. Moreover, all students were asked not to read the article content during the reading comprehension test. After finishing the test, the first phase of “Reading e-books without Mozart” was completed. After one week, the second phase also proceeded with a reading followed by a test, as in the first phase, except that Mozart’s K.488 music was playing during the reading task, with group I reading the “Rock Climbing” and group II reading “The Glacier.”

In order to let the students be able to listen the Mozart’s music without any interruptions, the door and windows of the classroom were closed during the related learning task, and each student’s personal computer and speaker were used to play K.488.
Results

Learning anxiety, reading rate and cognitive load

As seen in Table 3, the learning anxiety scores of both groups when Mozart was played were significantly higher than those seen for the silent task, and thus the students’ had lower learning anxiety when listening to this music. This result provides support for Hypothesis 1.

As for the reading rates, we found that those for the music task were higher than those for the silent task. The results thus indicate that listen to Mozart’s K.488 as background music can indeed affect students’ reading rates.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Silent task</th>
<th>Music task</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>34</td>
<td>23.50</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td>Group II</td>
<td>28</td>
<td>22.96</td>
<td>4.26</td>
<td></td>
</tr>
<tr>
<td>Reading rate (wpm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>34</td>
<td>282.58</td>
<td>10.09</td>
<td></td>
</tr>
<tr>
<td>Group II</td>
<td>28</td>
<td>282.21</td>
<td>10.14</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01; d = the effect size.

This study was to examine the correlation between students’ learning anxiety and reading rate using Pearson’s product-moment correlation coefficient, and data from both group I and II during the silent task was used. According to the results in Table 4, there was a significant, negative correlation between the learning anxiety and reading rate ($r = -.702, p < .01$) in group I, and a significant negative correlation between the learning anxiety and reading rate ($r = -.792, p < .01$) was also found in group II. Thus, the latter will increase if the former is reduced, and vice versa. This result provides support for Hypothesis 5.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Learning anxiety</th>
<th>Reading rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Learning anxiety</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Reading rate</td>
<td>-.792**</td>
</tr>
<tr>
<td>Group 2</td>
<td>Learning anxiety</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Reading rate</td>
<td>-.792**</td>
</tr>
</tbody>
</table>

Note. **p < .01.

As seen in Table 5, the intrinsic load scores were not statistically significant between silent task and music task of both groups, indicating that the different task had the same intrinsic load in this study. This result provides support for Hypothesis 2. However, the extraneous load scores when Mozart was played were significantly higher than those seen for the silent task, and thus the students’ had higher extraneous load when listening to this music. This result provides support for Hypothesis 3.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Silent task</th>
<th>Music task</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic load</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>34</td>
<td>4.41</td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>Group II</td>
<td>28</td>
<td>5.43</td>
<td>2.83</td>
<td></td>
</tr>
<tr>
<td>Extraneous load</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>34</td>
<td>5.74</td>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td>Group II</td>
<td>28</td>
<td>4.18</td>
<td>1.76</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05; ***p < .001; d = the effect size.
**Reading comprehension**

As seen in Table 5, we found the students’ reading comprehension in the music task was significantly better than that in the silent task. This shows that reading e-books while listening to Mozart can indeed improve students’ reading comprehension, which provides support for Hypothesis 4.

According to PIRLS (Ko, Chan & Chiu, 2013), the reading process can be divided into four phases, (1) focus and retrieve explicitly stated information, (2) make straightforward inferences, (3) interpret and integrate ideas and information, and (4) examine and evaluate content, language, and textual elements (MOE, 2011). Stages 1 and 2 belong to the direct process, while stages 3 and 4 are part of the interpretation process. We thus subdivided the reading comprehension performance of the tested students into the direct and interpretation components for further analysis and comparison. Since the total scores for the direct and interpretation processes of the two articles were different, the scores were standardized to avoid biased results.

In the direct process, we found that the students’ performance of two groups in the music task was significantly better than that in the silent task, again indicating that listen to Mozart’s K.488 as background music can help student’s reading. In contrast, the result for the students’ interpretation process performance of two groups in the music task was significantly lower than that in a silent task.

<table>
<thead>
<tr>
<th>Reading comprehension</th>
<th>Silent task</th>
<th>Music task</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Group I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>34</td>
<td>6.94</td>
<td>3.38</td>
<td>34</td>
</tr>
<tr>
<td>Direct process</td>
<td>34</td>
<td>38.23</td>
<td>19.54</td>
<td>34</td>
</tr>
<tr>
<td>Interpretation process</td>
<td>34</td>
<td>39.41</td>
<td>24.36</td>
<td>34</td>
</tr>
<tr>
<td>Group II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>28</td>
<td>7.75</td>
<td>2.19</td>
<td>28</td>
</tr>
<tr>
<td>Direct process</td>
<td>28</td>
<td>42.86</td>
<td>13.95</td>
<td>28</td>
</tr>
<tr>
<td>Interpretation process</td>
<td>28</td>
<td>43.57</td>
<td>18.10</td>
<td>28</td>
</tr>
</tbody>
</table>

*Note.* *p < .05; ***p < .001; d = the effect size.

**Students’ feedback from the questionnaires**

After the experimental activities, the students were asked the following two questions: First question “Did listening to Mozart help you to relax?” and second question “Did listening to Mozart help you to understand the content of the article?” There were 49 valid responses from all students for the first question, and 52 valid responses for the second. We invited two experts to code the open-ended questions. A Kappa coefficient was used to test the inter-rater reliability of the four sets of data from the coders, and this produced a Kappa coefficient of K = .795, p < .001. This shows that there was a significant correlation between the assessments made by two coders. Based on the results of experts to code the open-ended questions and descriptive statistics, a summary of the students’ answers is shown in Table 6.

<table>
<thead>
<tr>
<th>Type</th>
<th>Number/Total number of responses</th>
<th>Percentage</th>
<th>Reason</th>
<th>Number of students</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening to Mozart can help me relax.</td>
<td>46/49</td>
<td>93%</td>
<td>Music is nice</td>
<td>10</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>It makes me happy</td>
<td>9</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>I like music</td>
<td>9</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Relaxing</td>
<td>7</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Others</td>
<td>11</td>
<td>23%</td>
</tr>
<tr>
<td>Listening to Mozart cannot help me relax.</td>
<td>3/49</td>
<td>6%</td>
<td>Distraction</td>
<td>2</td>
<td>66%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Feel noisy</td>
<td>1</td>
<td>33%</td>
</tr>
<tr>
<td>Listening to Mozart can help me</td>
<td>49/52</td>
<td>94%</td>
<td>It is relaxing</td>
<td>22</td>
<td>44%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>It makes me happy</td>
<td>10</td>
<td>20%</td>
</tr>
</tbody>
</table>
understand the content of the article. | It helps me concentrate | 2 | 4% | Others | 15 | 30% | It has a negative influence on my reading | 2 | 66% | Feel nervous | 1 | 33%

For the first question “Did listening to Mozart help you to relax?” we found that 46 students (93%) felt that Mozart could help them relax, with the most common reasons being “Music is nice.” For the second question “Did listening to Mozart help you to understand the content of the article?” 49 students (94%) stated that listening to Mozart had a positive impact on their comprehension of the articles.

Discussion

How did Mozart affect learning anxiety, reading rate and reading comprehension?

Excessive anxiety can interfere with students’ cognitive processes, resulting in low performance (Wu & Lai, 2007) and poorer task efficiency (Gifford, & Marston, 1966; Seliger, 1972; Byrne & Eysenck, 1995). In contrast, Job and Dipamo (1991) also found that lower anxiety would result in better performance on a task. Previous studies have found that music can reduce anxiety and improve work performance (Dosseville, Laborde & Scelles, 2012; Lai, Chen, Chen, Chang, Peng & Chang, 2008), and that the slow tunes and classical music in particular have more positive impacts on reading comprehension (Thompson, Schellenberg & Letnic, 2012) and reading performance (Kallinen, 2002).

In this study, Mozart’s piece K.488 was used in e-book reading activities for elementary school students, and the results showed that it lowered their anxiety and raised their reading comprehension performance, which echoes the results of previous works (Dosseville, Laborde & Scelles, 2012; Kallinen, 2002). Sylvester (1995) argued that emotions can affect attention, and attention can affect learning, with negative emotions having negative effects. A number of related studies (Weiss & Cropanzano, 1996; Matuliauskaitė & Žemeckytė, 2011) also claimed that negative emotions can reduce task performance, such as learning achievement, as also found in the current study. Furthermore, we found that the group I’s average reading rate of 291.12 wpm in the music task was significantly higher than that of 282.58 wpm in the silent task, and group II’s average reading rate of 289.93 wpm in the music task was also significantly higher than that of 282.21 wpm in the silent task. Besides, we also found that learning anxiety and reading rate are negatively correlated with each other when students are reading e-books without background music. This finding is in line with previous studies (Gifford, & Marston, 1966; Seliger, 1972). Consequently, our findings are in line with those of related studies (Byrne & Eysenck, 1995; Gifford, & Marston, 1966) which reported that anxiety could adversely affect task efficiency and reading rate.

How did Mozart’s piece K/488 affect direct process and interpretation process performance?

In this study, we found that the direct process performance of students in the music task was better than that of students in the silent reading task. However, the interpretation process performance of students in the music task was lower than that of students in the silent reading task. This was probably because the capacity of Taiwanese fifth and sixth grade elementary school students is weaker with regard to the interpretation process than the direct process performance, based on PIRLS (Ko, Chan & Chiu, 2013). Compared with the direct process, the interpretation process is a higher level one that requires “interpretive integration” and “comparative assessment,” which often need more attentional resources (Cheng, 2006). In addition, fifth and sixth grade students are in the “Reading to learn” stage of reading development (Chall, 1983), and this focuses on cultivating the reading strategies that students need in order to acquire knowledge through reading. Although listening to Mozart can help reduce the anxiety of students, it did not seem to have a positive effect on the interpretation process. This can be explained by the Central Capacity Theory of the attention (Kahneman, 1973), which argues that a person’s attentional resources are limited, and multiple tasks will compete for the available attention (Kahneman, 1973). In addition to the above, Sweller et al. (2011) claimed that the cognitive system would fail if intrinsic and extraneous cognitive load exceeded the available resources of working memory. In this study, students’ extraneous load increased when they listened to Mozart music. The results show that the background music produced excess extraneous load in music task. Thus, since the students in the experiment were still in the stage of acquiring reading strategies, their abilities with regard to the higher level interpretation process required more...
training. Listening to Mozart might thus have distracted them, leaving too few attentional resources for effective learning to occur.

Conversely, the direct process, which involves tasks such as extracting information, and inference analysis, is a more fundamental learning capability that fifth and sixth grade elementary school students are already familiar with, and thus it requires relatively fewer attention resources. By reducing the anxiety, listening to Mozart may thus be able to promote the students’ direct process performance.

Research implications

The main contribution of this study was to discover that listening to Mozart, and especially his piece K.488, can provide fifth and sixth grade elementary school students with effective support, thus reducing their learning anxiety and improving their reading comprehension and direct process performance. However, and in line with the Central Capacity Theory (Kahneman, 1973), we also found that if the students were not familiar with the learning tasks or the strategies used then teachers should not impose too many external factors in the learning activities, such as music, in order to avoid overtaxing the students’ attentional resources and extraneous load. Therefore, if teachers want to play Mozart during reading activities, they should consider the difficulty of the learning task they have assigned, so that negative effects can be avoided.

Conclusion

This study has investigated how music, and specifically Mozart, can affect students’ learning anxiety, reading rate, cognitive load and reading comprehension when reading e-books, based on 62 senior-grade elementary school students. Based on our findings, we found that Mozart had positive effects on the students’ learning anxiety, reading rate, reading comprehension, and the direct process performance in the e-book reading tasks. However, it had a negative impact on the students’ interpretation process performance and extraneous load. We inferred that listening music might distract the students as it could take part of resources for attention on the reading task. Therefore, teachers should choose reading tasks with an appropriate difficulty if they wish to play Mozart in class.

It is worth mentioning that the Taiwanese government has initiated a series of pilot projects with regard to adopting e-books in primary and secondary school teaching activities (MOE, 2010). Unlike traditional paper books, e-books can also present multimedia content (Huang & Liang, 2015), and thus they can easily be used to play music, such as Mozart, in order to promote reading effectiveness.

This study is based on data obtained using a subjective learning anxiety scale. In order to explore the related issues more objectively, future research could consider using an electroencephalogram to measure students’ learning anxiety.

Acknowledgements

This work was supported by the Ministry of Science and Technology, ROC, under Contracts No. MOST 103-2511-S-006-002-MY3, and MOST 103-2511-S-006-007-MY3.

References


110


Ministry of Education (MOE). (2007). 國民中小學提升閱讀計畫 [The Plan of enhance the reading for Primary and secondary schools]. Retrieved from http://reading.kh.edu.tw/plan/upload/%E6%95%99%E8%82%B2%E9%83%A8%E6%82%85%E8%AE%80101%E8%A8%88%E7%95%AB.pdf


Ken Nee Chee1*, Noraffandy Yahaya1, Nor Hasniza Ibrahim1 and Mohamed Noor Hasan2

1Department of Educational Science, Mathematics and Creative Multimedia, Faculty of Education, Universiti Teknologi Malaysia // 2Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia // knchee2@live.utm.my // p-afandy@utm.my // p-norhaniza@utm.my // mnoor@utm.my

*Corresponding author

(Submitted October 20, 2015; Revised March 22, 2016; Accepted June 2, 2016)

ABSTRACT

This study examined the longitudinal trends of mobile learning (M-Learning) research using text mining techniques in a more comprehensive manner. One hundred and forty four (144) refereed journal articles were retrieved and analyzed from the Social Science Citation Index database selected from top six major educational technology-based learning journals based on Google Scholar metrics in the period from January, 2010 to December, 2015. Content analysis was implemented for further analysis based on (a) category of research purpose, (b) learning domain, (c) sample group, (d) device used, (e) research design, (f) educational contexts (i.e., formal learning and informal learning), (g) learning outcome (i.e., positive, negative and neutral), (h) periodic journal, (i) country, and (j) publisher. This review study of M-Learning presents findings, which can become a layover platform and guidance for researcher, educators, policy maker or even journal publisher for future research or reference in the realm of M-Learning regarding the latest trends.

Keywords

Mobile learning, M-Learning, Research trends

Introduction

With the advent of mobile technologies, new paradigm of teaching and learning with technology aid had been emerged, that is mobile learning (M-Learning). Mobile technologies purvey opportunities to hold new and interesting methods of teaching and learning, both beyond and inside the classroom. Apropos to the teaching-learning process, the integration of mobile devices into educational context has considerable benefits and profoundly potential which consistence with Valero et al. (2012) who claimed that the technological features of M-Learning are portability, immediacy, connectivity, ubiquity and adaptability. It enables collaboration among pupils, knowledge creation, information searching and improved interaction and communication between teacher and student. Moreover, it facilitates access to learning anytime and anywhere by enabling connectivity and the employ of multiple apps for educational purposes (Fundación Telefónica, 2013). In short, M-Learning has been recognized as one of the most influential technologies for education (Johnson, Adams, & Cummins, 2012).

Therefore, this paper intends to provide insights into the research trends and issues in the studies of M-Learning through content analysis of selected journals from January, 2010 to December, 2015, covering six major journals: (1) Computer & Education (C&E), (2) British Journal of Educational Technology (BJET), (3) Educational Technology & Society (JETS), (4) Journal of Computer Assisted Learning (JCAL), (5) The Internet and Higher Education (IHE) and (6) The International Review of Research in Open and Distance Learning (IRODL).

This study reported herein, investigated longitudinal trends of M-Learning research with text mining techniques. In sum, this study systematically reviews and synthesizes the relevant literature through a meta-analysis (Creswell, 2002, pp. 351-353) to provide a more comprehensive analysis of previous studies.

Specifically, the present study poses the four research questions:

- What are the sources of the article regarding periodic journal, publisher and country that were related to M-Learning that were published in these selected journals from 2010 – 2015?
- What are the main research purposes, sampling and outcome/conclusion that was related to M-Learning that were published in these selected journals from 2010 – 2015?
- What are the learning domains, device used, and educational context related to M-Learning that were published in these selected journals from 2010 – 2015?
- What types of research design have been applied in article research of M-Learning that were published in these selected journals from 2010 – 2015?
Literature review

Definition of M-Learning

M-Learning, which means learning through mobile devices (such as smart mobile phones and tablet PCs), is changing the educational environment by offering learners the opportunity to engage in asynchronous, ubiquitous instruction (Hyman et al., 2014). M-learning is a teaching method that has the intersection between mobile computing and e-learning (Quinn, 2000; Keengwe, 2014) that integrates several software and firmware technology in multimedia applications (Lavin-Mera et al., 2008) which facilitate learning through a variety of wireless mobile devices (Kukulska-Hulme, 2005; Stevens & Kitchenham, 2011) using wireless networks (WiFi) or broadband services (Caudill, 2007) without limit in terms of location or time. (Kukulska-Hulme, 2005; Hussin et al., 2012; Quinn, 2000). Furthermore, Keegan (2002) contemplates the possibility of M-Learning as a harbinger of the future of learning.

M-Learning research

The use of mobile devices for educational purposes, recognized as M-Learning has gained substantial attention from researchers in the technology-enhanced learning discipline. Recent research findings on using mobile devices in different learning environments have exemplified their ability to effectively enhance students’ learning knowledge. Understanding and experience in divergent subject areas such as science (Looi et al., 2011; Hwang Wu, & Ke, 2011; Ahmed & Parsons, 2013), mathematics (Huang et al., 2012; Mahamad et al., 2010; Lan et al., 2010), language and art (Yu et al., 2013; Martin & Ertzberger, 2013), social science (Shih et al., 2010), engineering (Yang et al., 2013) and others. This promising role in education can tremendously be noticeable within the informal and formal learning context, such as guiding an interactive tour with museum visits (Sung et al., 2010; Hou et al., 2014) facilitating knowledge acquisition in field trips (Menkhoff & Bengtsson, 2012), game-based learning (Young et al., 2012), in-class collaboration learning (Echeverria et al., 2011). Nevertheless, there is always a contrasting scenario in every context, including M-Learning as Chu (2014) argued that the performance of students, known to be “effective,” might be disappointing or may even negatively affect the students’ learning achievements if without proper treatment employed.

Previous review paper on M-Learning

In recent years, there were three literature reviews with high citation as in December, 2015 in Google scholar studied on research trends in M-Learning. Literature review paper with the title, “Examining M-Learning trends 2003–2008: a categorical meta-trend analysis using text mining techniques” which written by Hung and Zhang (2012) and cited 45 times according to Google Scholar, used text mining techniques to investigate research trends in 144 academic articles based on five journal include Lecture Notes in Computer Science (LNCS), JETS, JCAL, C&E, and International Journal of Engineering Education on mobile learning (IJEEML) from 2003 to 2008 taken from the SCI/SSCI database. In general, they investigated publication date, publication category, taxonomy, article clusters, and country, university and journal of origin. Results showed that articles on M-Learning increased from 8 in 2003 to 36 in 2008; the most popular domains in M-Learning studies are effectiveness, evaluation, and personalized systems and studies on strategies and frameworks are more likely to be published. Apart from that, they found that Taiwan is the most contributing country and university regarding journal publications on M-Learning.

Another review paper entitled “Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010” which written by Hwang and Tsai (2011) and cited 121 times according to google scholar, reviews the advancement of mobile and ubiquitous learning research from 2001 to 2010 by selecting 154 articles on mobile and ubiquitous learning based on the articles published in six major SSCI journals included BJET, C&E, JETS, Educational Technology Research & Development (ETRD), JCAL and Innovations in Education and Teaching International (IETI). It is found that the number of articles has significantly increased during the past 10 years; moreover, researchers from other countries have contributed to the related field in recent years. Scope of the review included a number of articles published, research sample groups selected, research learning domains, and country of origin. They found out that research in mobile and ubiquitous learning increase drastically in number between 2006 and 2010; higher education students were the most frequent research sample, followed by elementary school students and high school students; most studies did not explicitly focus on any particular learning domain but rather investigated the motivation, perceptions and attitudes of students toward mobile and ubiquitous learning, along with course-orientation for engineering.
(including computers), language and art, and science; and most articles were contributed from US-based authors, followed by authors in the UK and Taiwan for the first five years and it was vice versa for the another second five years.

Following these two literature reviews-based studies, another review paper entitled “Review of trends from M-Learning studies: A meta-analysis” which written by Wu et al. (2012) and cited 162 times according to Google scholar comes about to step into the breach since there were issues that still needed to be examined from other directions such as the distribution of research purposes. This study takes a meta-analysis approach to systematically review the literature of 144 studies based on the articles published in six major SSCI journals included JCAL, Computer in Human Behavior (CHB), BJET, JETS, and IRODL from 2003 to 2010. Major findings include that most studies of M-Learning focus on effectiveness, followed by M-Learning system design, and surveys and experiments were used as the primary research methods. Apart from that, mobile phones and PDAs are currently the most widely used devices for M-Learning, but these may be displaced by emerging technologies. Moreover, most M-Learning studies feature positive outcomes and M-Learning is more prevalent at higher education institutions, followed by elementary schools. In addition, the most highly-cited articles are found to focus on M-Learning system design, followed by system effectiveness.

Apart from the above mentioned on high cited review paper, there were another review paper that ought to be included in this section which is review paper entitled “Applications, impacts and trends of mobile technology-enhanced learning: a review of 2008–2012 publications in selected SSCI journals” which written by Hwang and Wu (2014) and cited 26 times according to Google scholar, reviews the 214 publications from 2008 to 2012 in seven well-known SSCI journals of technology-enhanced learning included C&E, JETS, Educational Technology Research and Development (ETRD), IETI, BJET, JCAL and Interactive Learning Environments (ILE) as to examine on the applications and impacts of mobile technology-enhanced learning. It is found that M-Learning is promising in improving students’ learning achievements, motivations and interests with proper use of mobile technologies and education design together with proper support and strategy; top four applications were language learning, environmental and ecological education, engineering and computer education and historical and cultural education; most of the applications were conducted both indoor and outdoor activities indoors, followed by indoor and then outdoor; smartphones and followed by Personal Digital Assistants (PDAs) are the most frequently used M-Learning devices, and only then tablet PCs, but smartphones and tablet PCs had replaced on the use of PDAs in educational settings which started from 2011 and 2012; mobile technologies have been increasingly applied to formal and informal.

After all, recent literature review paper seems to be filling in the breach of previous review papers, which were incomplete and act as complementary. This study adopts a meta-analysis method in examining trends in M-Learning studies in term of the various criteria across years in the period under review comprehensively all in one as to refine and update with the most present M-Learning trend. These findings may provide insights for researchers and educators, even policy makers into research trends in M-Learning.

Method

Data sources and search strategies

This study examines the M-Learning papers published in the SSCI database from 2010 to 2015. Top six major educational technology-based learning journals were selected to analyze the research trends, including the (1) C&E, (2) BJET, (3) JETS, (4) JCAL, (5) IHE and (6) IRODL. These journals are widely accessed with high impact factors based on top publication reports released by the Google Scholar metrics. The thorough and plenary searching were through manual electronic searches of the following databases: Science Direct for journal (1) and (5), ProQuest for journal (3) and (6) and Wiley Online Library for journal (2) and (4).

Two researchers who have had years of experience carrying out studies in this area were asked to filter the M-Learning studies from the 1338 papers published by these six journals (378 from BJET, 61 from JCAL, 492 from C&E, 70 from IHE, 243 from JETS, and 94 from IRODL) from 2010 to 2015. Only papers that were identified as being of the type “articles” in the SSCI were considered; that is, publications such as “book reviews,” “letters,” “colloquium,” “conference paper,” “workshop paper,” “presentation paper,” “book chapter,” “proceeding,” “thesis,” “dissertation” and “editorial materials” were all excluded from this study. We intend to include all of the papers published in these journals about Mobile Learning and M-Learning without utilizing other filtering criteria. It is expected that such a review can provide a more thorough view of M-Learning research. To be more precise in selecting the M-Learning articles from the candidate pool, the articles selected by
the two researchers were compared to see if there were inconsistent selections, and if so, these selections were shown to the researchers for further discussion. A total of 144 studies concerning M-Learning were selected after two iterations of filtering the papers and discussing on the inconsistency of decisions.

Data coding and analysis

Ten features related to the quality of study research methodology were coded, including (a) category of research purpose, (b) learning domain, (c) sample group, (d) device used, (e) research design, (f) educational contexts (i.e., formal learning and informal learning), (g) learning outcome (i.e., positive, negative and neutral), (h) periodic journal, (i) country, and (j) publisher.

This study uses the methodology of content analysis to analyze trends and issue about M-Learning. Stemler (2001) confirmed that content analysis indeed is a powerful method for examining trends and patterns in documents. It is also a useful technique to discover and describe the focus of individual, group, institutional or social attention (Weber, 1990). By conducting a content analysis from the 144 selected journals in the timeframe of 2010 to 2015, this study will look out for issues and trend that underlies the studies of M-Learning currently. Besides that, this study cross-examines papers related to M-Learning; published in six selected journals from 2010 to 2015. Three databases were chosen for the cross-examine purpose. The different databases were chosen due to the availability of certain journals and accessibility of the abstract and full text of the selected articles. The databases were; ProQuest Education Journals, Science Direct and Wiley Online Library. Google scholar as a search engine was also used for the purpose above.

The first procedure in conducting this research is setting three items to search for the related articles in all databases above. They are; (1) Selected Journal Name for Journal Name, Publication Title or Journal Title column, (2) mobile learning for Topic or Title column and (3) 2010-2015 in Time span, Year or Coverage column. This step is important to ensure standardization in order to search the related articles in spite of the different interface between all databases.

There were 162 articles have been identified from the first procedure. The next procedure consists of further comprehensive review, which needs the researchers to examine 162 articles carefully to determine the articles which is related to M-Learning. Finally, a total of 144 articles were selected for the analysis.

Trend analysis

Trend analysis of an article can show the periodic discussion taking place in a knowledge discipline (Erford et al., 2010). In the analysis of trend and frequency, justification for selection of articles is found in the BJET, JETS, C&E, ETS, IHE and IRODL only.

Content analysis

Based on content analysis or the process of summarizing and reporting of written data (Hsieh & Shannon, 2005). The research topics in the articles selected for analysis were categorized involves counting and comparisons according to key words in the given abstracts and content, issues discussed as well as research scope followed by the interpretation of the underlying context. Throughout the data analysis carried out, each category identified was further clarified using thematic analysis.

Result

Research question 1

Trend of periodic journal contributing to M-Learning field across the years

As depicted in Figure1 and Table 1, out of top six major educational technology-based learning journals that were selected to analyze the research trends, which including the C&E, BJET, JETS, JCAL, IHE and IRODL, it is obvious that JETS (26.39%) were the most contributing journal towards M-Learning field till peak in 2014 then drop abruptly in 2015, which causes BJET (27.78%) leads in front for the six year period due to its sudden
increment in 2015. This rank is followed by C&E (23.61%), IRODL (13.89%), JCAL (6.26%) and IHE (2.08%). BJET and C&E saw a dramatically growth between the six year period. The rest fluctuated unevenly over the years. Overall, there is a tremendous increment in total from year to year except in 2011.

### Table 1. Distribution of M-Learning studies by periodical journal across years from 2010 to 2015

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C&amp;E</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>15</td>
<td>34</td>
</tr>
<tr>
<td>BJET</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>11</td>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td>JETS</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>14</td>
<td>2</td>
<td>38</td>
</tr>
<tr>
<td>JCAL</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>IHE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>IRODL</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Sum</td>
<td>17</td>
<td>14</td>
<td>20</td>
<td>22</td>
<td>33</td>
<td>38</td>
<td>144</td>
</tr>
</tbody>
</table>

**Figure 1.** Distribution of M-Learning studies by periodical journal across years from 2010 to 2015

**Trend of publishers that contributing to M-Learning field across the years**

Two periodic journals embodied in each publisher database. Science Direct consist of journal C&E and IHE, while ProQuest comprises of journal JETS and IRODL and whereas Wiley Online Library contain journal BJET and JCAL. Since database sources are linked to the periodic journal, it is acceptable that ProQuest headed up all the rest and so violently increase in between 2012 to 2014 due to the proliferation of total number of journal JETS and IRODL followed by Wiley Online Library and Science Direct as illustrated in Figure 2. Apart from that, Wiley Online Library rises gradually along these years while Science Direct shown considerable fall in 2014 and then increase tremendously in 2015 due to the proliferation of journal C&E in 2015.

**Figure 2.** Distribution of M-Learning studies by publisher database across years from 2010 to 2015

**Trend of countries that contributing to M-Learning field across the years**

As indicated in Figure 3, it is perceivable that more country has contributed their research on M-Learning as there are new emerging country like China, Malaysia, Sri Lanka, Pakistan, Iran and several more. This may due to the existence of awareness on the significance of M-Learning as a new and trendy teaching and learning paradigm in this advent of the technology era. Conspicuously, Taiwan is the most dominance country contributing to M-Learning research with the total up across the years at 25.30%, followed by USA (15.06%),
United Kingdom (7.23%), Singapore (6.63%), Turkey (6.02%), Canada (5.42%), and others country with percentage less than five percent.

**Research question 2**

*Trend of issue category regarding M-Learning across the years*

Articles were categorized into one of four categories according to its research purpose: (1) evaluating the effects of M-Learning, (2) designing a mobile system for learning, (3) elicit perceptions of M-Learning, (4) review on M-Learning or (5) evaluate or explore the factor towards M-Learning. As delineated in Figure 4, evaluating the effects of M-Learning was the most common research purpose (52.53%) which mainly covered large portion of the stacked area line chart across years, followed by review on M-Learning (17.09%), evaluate or explore the factor towards M-Learning (15.82%), elicit perceptions of M-Learning (7.59%) and designing a mobile system for learning (6.96%). Category of evaluating the effects of M-Learning start to increase progressively in 2012 till 2015. For the category of evaluating or explore the factor towards M-Learning, it showed gradually rising along these years. Nevertheless, the rest categories had shown fluctuation along these years.

**Trend of sampling taken regarding M-Learning across the years**

By exclusion from review paper, Figure 5 shows that M-Learning research mainly focuses on higher education institution (36.17%), followed by not specific (35.11%), elementary or primary school (21.28%), High or Secondary School (6.38%) and the rest were working adult. There was a sharp shoot up in 2014 for the number
of articles using higher education as sample institution while the rest was showing up and down unstably. Besides, Fig. 6 indicates that higher education student leads the trend (50.75%), followed by elementary or primary school student (19.40%), elementary or primary school teacher (13.43%), high or secondary school student (7.46%) and lastly tailed by higher education instructor (1.49%). A number of articles that were utilizing higher education students as the sample are topping all others sample, but it's shown unstably fluctuate as others sample except in year 2015.

Figure 5. Distribution of M-Learning studies by sample institution across years from 2010 to 2015

Figure 6. Distribution of M-Learning studies by sample individual across years from 2010 to 2015

Trend of outcome/conclusion resulting in M-Learning across the years

Despite of irrelevant outcome (37.66%) synthesized by others from the evaluating effect purpose, Figure 7 indicates that 52.60% of studies reported positive research outcomes, while only 6.49% and 3.25% respectively reported neutral and negative outcomes generated from the journal with evaluating effect purpose. All the outcomes showed a steady increase along the period.
Research question 3

Trend of learning domain regarding M-Learning across the years

Regardless of journal without specific on learning domain with reach 53.06%, a majority of published M-Learning studies focused on two subject areas: Science (12.24%) and Language and Art (12.93%). Additional studies were conducted in fields like Social Science (8.16%), others (6.80%), Engineering (4.08%), and Mathematics (2.72%). Despite of that, Science, and Language and Art peaked in 2015 although all categories shown up down pattern.

Trend of device used regarding M-Learning across the years

Mobile phone in this study referred to the basic cell phone without the function that exist in a smartphone, which included 3G/4G, or Wi-Fi connection. Term of smartphone used in this study is a general term without specifying in android or iOS as a platform. In spite of journal without specific stated device used (57.02%), Fig. 9 indicates that, among the 144 studies, smart phone was most commonly used for M-Learning (14.09%), followed by PDAs (8.05%), mobile phone (7.38%), tablet (5.37%), iPad (4.70%), iPhone (2.68%), and iPod (1.34%) in total. All the line moves unstable along the period as the preferred device used in M-Learning research, however, it can be observed that PDA has shown a sharp drop starting in the year 2013.
Trend of educational context regarding M-Learning across the years

As can be seen in Figure 10, informal learning (11.11%) was predominant in the M-Learning studies compared to formal (8.33%) and a combination of both (6.25%). Formal line and informal line showed gradual increase along these years while a combination of formal and informal line shown up and down.

Research question 4

Trend of research design regarding M-Learning across the years

Quantitative approach (47.92%) is the most employed research designs for M-Learning research studies, followed by a mixed method (18.75%) and Qualitative (14.58%) as depicted in Figure 11. Out of 144 articles analyzed, there were 18.75% articles with no specific approach due to the existence of the review paper. There was a dramatically shot up shown by the quantitative design line in 2012, whereas others shown unstable rise and fall across the years.
Discussion

Based on several studies selected and merely four literature review papers as reference, this review paper can be produced in a more detailed and refined even up-to-dated all in one version. This is because in the previous three review papers on M-Learning recently authored by Hwang and Tsai (2011), Hung and Zhang (2012), Wu et al. (2012), and Hwang and Wu (2014) are just a compliment to each other with the older version of the newborn.

Wu et al. (2012) reproved on the previous two is still incomplete in their criteria and the topic being further explored from different directions. This study imparts comprehensive results and new findings. For example, this research found that most M-Learning research paper could be obtained in certain journals like JETS, BJET and C&E whereas in a database like ProQuest based on the frequency count and even country that most M-Learning research is derived from Taiwan followed by the USA which in line with Hung and Zhang (2012). There were more findings that will further describe at below.

Taiwan is the most dominance country contributing to M-Learning research

As claimed by Hung and Zhang (2012) in the research period, and Hwang and Tsai (2011) in the second half period, Taiwan has become the top country regarding M-Learning research which corresponds with finding in this paper.

BJET and JETS are the most Periodic Journal while ProQuest is the most Publisher Contributing to M-Learning Field

From the result, it is affirmed that BJET and JETS are the most Periodic Journal while ProQuest is the most Publisher Contributing to M-Learning Field.

Most studies of M-Learning focus on effectiveness, followed by M-Learning review

Out of the 144 studies, 52.53% took evaluating the effectiveness of M-Learning as the main research purpose as depicted in Figure 4. This focus on effectiveness evaluation is in line with Wu et al. (2012), and Hung and Zhang (2012). The second-most frequently cited research purpose was M-Learning review, which is also a new finding which is contrary to Wu et al. (2012), and Hung and Zhang (2012). More importantly, we found that the number of studies devoted to all M-Learning research increased over time, which supported by Hwang and Tsai (2011) and Hung and Zhang (2012). This may be due to the advent of mobile technology and the enormous advantages that bring along with and mean that the trends are still keep increasing till to date.
Most M-Learning studies took samples from a higher education institution, followed by the elementary or primary school

As seen in Figure 5, the result which consistent with Wu et al. (2012) revealed that higher education institution is the main sampling pool regarding to M-Learning research may be due to the convenience factor. This is because researchers mostly were originated in university or college. Primary school is the next most sampling taken from. Reason behind this need to be justified in the future research. As a result, there is tremendous room for research to be carried out for others sample such as secondary or high school and working adult.

Most M-Learning studies took higher education students as sample, followed by elementary or primary school student

As shown in Figure 6, the result which consistent with Hwang and Tsai (2011) revealed that higher education student is the main sampling pool regarding to M-Learning research may be due to the convenience factor, the same reason as a sampling institution above. Again, this is because researchers mostly were originated in university or college. The primary school student is the next most sampling taken from. Reason behind this need to be justified in the future research. As a result, there are tremendous room for research to be carried out for others sample such as working adult, primary or elementary school teacher, secondary or high school teacher and student.

Most M-Learning studies feature positive outcomes

Figure 7 shows that most of the 144 M-Learning studies present positive outcomes. This finding corresponds to the finding from (Wu et al., 2012). Neutral outcome ranked next and negative outcome ranked the least.

M-Learning most frequently support learning in the Language and Art, followed by Science

Figure 8 illustrates that studies on M-Learning in educational contexts, most frequently focus on use in supporting subject Language and Art, followed by the Science, Social Science, others, Engineering and Mathematics. In terms of M-Learning activity in various sub-disciplines, our findings partially support those of Wu et al. (2012), and Hwang and Wu (2014) but fully support to Hwang and Tsai (2011). For instance, Wu et al. (2012), and Hwang and Wu (2014) showed M-Learning was often used in language courses. Profoundly, the present study found that M-Learning is also widely used in courses related to Science, Social Science, engineering and others but considerably less in other courses such as Mathematics. Nevertheless, there is scarcity of M-Learning research in the related fields should be emphasized in the future research conducted as to fill in the gap.

Smartphone currently is the most widely used devices for M-Learning

The type of devices that were used in the context of M-Learning is influenced by the mobile consumer preference. Figure 9 indicates that smartphones are most widely used as teaching and learning tool in educational contexts corresponds with the Mobile Consumer Report (Nielson, 2013) which stated that smartphone owners may be the majority of mobile users in countries like the US and UK PDAs ranked second as it has been used as learning tools a decade ago and thus supporting the result from Wu et al. (2012), and Hwang and Wu (2014) but it is shown a sharp drop starting in year 2013 due to displacement of smartphones and tablet PCs as emerging technologies over the use of PDAs in educational settings consistent with Wu et al. (2012), and Hwang and Wu (2014).

McQuiggan et al. (2015) affirmed that it is widely predicted that mobile devices are the wave of the foreseeable future in educational technology. Thus, through the advancement of technology, the invention of new mobile devices will never come to an end and it will be applied to the educational context if its efficacy towards the field. This is supported when Martin et al. (2011) used predictions from 2004 to 2010 (i.e., from seven Horizon Reports), which cover the period 2004–2014, to analyze the technologies that have impacted education in the past or are likely to have an impact in the future. Horizon report 2007 predicted that the use of mobile phones in M-Learning, particularly in higher education, would increase dramatically after 2009, which corresponds with our findings.
Informal learning is the most preferred approach carried out along with M-Learning

As depicted in Figure 10, informal learning dominates the M-Learning context which in line with Traxler (2007) claimed that M-Learning definition can emphasize those unique attributes that position it within informal learning, rather than formal.

Most M-Learning studies adopted quantitative method as the primary research design

Figure 11 shows that, among the 144 studies, quantitative approaches were favored over mixed method approach and qualitative approaches. This finding corresponds with finding from Wu et al. (2012).

Conclusions

Three previous literature review-based studies on the use of M-Learning in academic contexts provided valuable insights, but they were just a compliment amongst them to cover up their incompleteness. This study was conducted a systematic meta-analysis to provide more comprehensive analysis of past studies, refined on previous review studies, and discusses the implications of new findings.

The current study presents nine new findings: (1) Taiwan is the most dominance country contributing to M-Learning research. (2) BJET and JETS are the most periodic journal while ProQuest is the most publisher, contributing to the M-Learning field. (3) Most studies of M-Learning focus on effectiveness, followed by M-Learning review. (4) Most M-Learning studies took sample from higher education institution, followed by elementary or primary school. (5) Most M-Learning studies took higher education students as sample, followed by elementary or primary school student. (6) Most M-Learning studies feature positive outcomes. (7) M-Learning most frequently supports learning in the Language and Art, followed by Science. (8) Smartphone currently is the most widely used devices for M-Learning. (9) Informal learning is the most preferred approach carried out along with M-Learning. (10) Most M-Learning studies adopted quantitative method as the primary research design. As a conclusion, this study of issues in M-Learning presents findings, which can become a layover platform and guidance for researchers, educators, policy makers or even journal publishers for future research or reference in the realm of M-Learning.

Implications for research and practice

The findings of this study contribute to an in-depth understanding of M-Learning, by providing a broad and a longitudinal overview of reputable publications according to Google Scholar metrics. It provides a quick, comprehensive overview for scholars interested in publications on M-Learning. For instance, researchers know which journal to be targeted on when M-Learning take its place. It has also identified the topics and areas that have been studied more intensively regarding M-Learning. Furthermore, the findings suggest topics and areas needing additional research to fill in the gap. Thus, researchers should pay more attention to the gap that is a scarcity of research and development of M-Learning in order to synthesize knowledge in the field.

As an emerging research method, text mining enables researchers to obtain summative information in virtually any given field. This study illustrates the power and potential of text mining techniques to discover research patterns, themes, and trends. These techniques enable scholars to pay more attention to data interpretation and pattern analysis, comparing to traditional information processing or data (content) analysis.

For government policy makers, the findings will provide supporting information to enhance understanding of research strengths and weaknesses, which in turn can influence decision-making and policy change towards the advancement in educational discipline.

For researchers, this finding will give a bigger picture on how importance of M-Learning as it gains more and more attention from all over the world due to the proliferation of country that have embarked on this new and trendy paradigm of teaching and learning method in education fields. Researchers and educators will ascertain on where to find about and target on M-Learning research with remarkable quantity and quality articles.

For journal publishers, this finding will notify on the statistics about M-Learning research published in their journal or even their database so that call for paper on M-Learning will be ushered in as to lure more papers
regarding M-Learning into particular journal publisher if it is necessary and create a healthy competition in the publication battlefield.

Limitation of the study

The results and conclusion are limited and not intended to be exclusive. SSCI journals adopt stringent journal reviewing criteria. Articles might take 2 years from submission to publication. In addition, the SSCI database does not collect conference proceedings in education. Therefore, the findings in this study may not reflect the most recent research trends.

This study used only two search terms to analyze M-Learning publications from the beginning of 2010 to year end of 2015 collected in the SSCI databases at that time. Future studies with greater resources, using more search terms, are needed to expand these findings.

References


Document Reproduction Service No. ED 472 435)
Keengwe, J., & Maxfield, M. B. (2014). Advancing higher education with mobile learning technologies: Cases, trends, and
inquiry-based methods. Hershey, PA: IGI Global.
contexts. In Digital Games and Intelligent Toys Based Education, 2008 Second IEEE International Conference on (pp.
44-51). doi:10.1109/DIGITEL.2008.21
experience for primary school students: A Study of learning effectiveness. Journal of Computer Assisted Learning, 27(3),
269-287.
Martin, F., & Ertzberger, J. (2013). Here and now mobile learning: An Experimental study on the use of mobile
Menkoff, T., & Bengtsson, M. L. (2012). Engaging students in higher education through mobile learning: Lessons learnt in a
Shih, J. L., Chuang, C. W., & Hwang, G. J. (2010). An Inquiry-based mobile learning approach to enhancing social science
interdisciplinary mobile learning: Delivering information to students (pp. 1-25). doi:10.4018/978-1-60960-511-7.ch001
Traxler, J. (2007). Defining, discussing and evaluating mobile learning: The Moving finger writes and having writ... The
International Review of Research in Open and Distributed Learning, 8(2). doi:10.19173/irrodl.v8i2.346
Yang, G., Chen, N. S., Sutinen, E., Anderson, T., & Wen, D. (2013). The Effectiveness of automatic text summarization in
princess is in another castle a review of trends in serious gaming for education. Review of educational research, 82(1), 61-89.
Yu, Y. S., Lin, Y. Y., Huang, Y. L., & Hsieh, W. H. (2013). The Evaluation of use the mobile phone learning English in
Taiwan. International Journal of Information and Education Technology, 3(2), 189-191.
The Effectiveness of Using Cloud-Based Cross-Device IRS to Support Classical Chinese Learning

Yi-Hsuan Wang

Department of educational technology, Tamkang University, Taiwan // annywang12345@hotmail.com

(Submitted October 31, 2015; Revised April 28, 2016; Accepted August 3, 2016)

ABSTRACT

The purpose of the present study was to examine the effects of integrating a cloud-based cross-device interactive response system (CCIRS) on enhancing students’ classical Chinese learning. The system is a cloud-based IRS system which provides instructors and learners with an environment in which to achieve immediate interactive learning and discussion in the classroom. A quasi-experimental design was employed in which the experimental group (E.G.) learned classical Chinese with the system, while the control group (C.G.) followed their original learning method. The results revealed that the novice and medium-achievement learners in the E.G. performed significantly better than other E.G. students, and most students as well as the instructor gave positive feedback regarding the use of the system for course learning. In sum, CCIRS is an easy-to-use learning trigger that encourages students to participate in activities, arouses course discussion, and helps to achieve students’ social and self-directed learning. The study concludes that the idea of “bring your own device” could be implemented with this system, while integrating educational factors such as game-based elements and competitive activities into the response system could reinforce flipped classroom learning.

Keywords

IRS, Cloud-based application, Chinese learning

Introduction

How to host and create interaction among students and teachers in class has always been an important challenge for course instructors (Pond, 2010; Wang, 2015). For the past few decades, technology-enhanced learning tools such as the Clicker technology (Lin, Liu, & Chu, 2011), the Interactive Response System (IRS) (Pond, 2010) or mobile-based applications for interactive response learning (Chuang, 2015; Stowell, 2014) have helped teachers overcome these challenges and have promoted students’ engagement and learning interaction in class. The basic components of the above-mentioned technology are hand-held transmitters and receivers that are used for delivering and receiving responses between teachers and learners, and software installed on teachers’ computers to present students’ responses to questions (Pond, 2010). The use of Clickers, IRS or interactive applications has the advantage of motivating participants to become more engaged than when such systems are not used in the classroom (Heaslip, Donovan, & Cullen, 2013) while real-time interactive systems help learners achieve better learning performance (Pond, 2010). Research has also indicated that adopting a student response system in the classroom benefits learners’ engagement, motivation and learning (Wang, 2015), and the use of IRS can play a key role in delivering pedagogical outcomes that motivate students to become more attentive and involved in class (Heaslip, Donovan, & Cullen, 2013). Nevertheless, there are some limitations to the above-mentioned educational technology tools (Clicker, IRS and App-based systems) including the effort required to administer the system environment (Barber et al., 2007), the cost of special devices such as clickers (Boatright-Hortwitz, 2009), the price of installing specific interactive software (Keough, 2012), the time cost of setting up the IRS environment in the limited class time (Keough, 2012), and teachers’ need for particular technology skills training before adopting the technology in their classrooms for both software and hardware installation (Wang, 2015). However, the most important problem is that each brand of clicker or mobile-based IRS application only accepts particular information sending devices (Stowell, 2014), meaning that if an IRS App is developed for smartphones, learners could not send information from computer-based equipment, or if the IRS App is developed for the iOS system, learners with Android phones could not participate in the IRS activity.

Chinese teaching in school includes modern and classical Chinese learning. The purpose of Modern Chinese learning focuses on language skills for daily use, while the Ministry of Education in Taiwan (Ministry of Education, 2013) announced that the purpose of classical Chinese learning is to promote students’ abilities of classical Chinese culture learning, reading comprehension, appreciation and learning motivation. Research has indicated that through the comprehension and memorization of classical Chinese, learners improve their capability of reading and writing modern Chinese (Chen, 2003). However, the wording and sentence construction of classical Chinese are much more complicated than in modern Chinese, so how to engage senior high students’ motivation for classical Chinese learning in lecture classes, and further, how to encourage students to achieve self-directed learning are still issues that need to be explored (Chiang, 2014).
Research purpose and questions

Several studies have investigated the potential use of interactive learning technology in fostering learning (Pond, 2010, Wang, 2015); however, there are some limitations of IRS, especially for cross-device usage, and the setup fees of the hardware and software system environments are high (Barber et al., 2007). To date, not many studies have made attempts to address the effectiveness of integrating cloud-based techniques to assist learning. Consequently, the purpose of this study was to integrate a cloud-based and cross-device interactive response system into a formal classroom, and to explore whether the assistance of this advanced learning technology could promote students’ learning performance, and further, whether it could achieve self-directed learning purposes. Based on the above rationale, the following research questions were investigated: (1) Could CCIRS promote learners’ classical Chinese learning performance? (2) How do the learners and instructor perceive the use of CCIRS for conducting learning activities? and (3) Does the integration of CCIRS form a different classroom teaching and learning pedagogy?

Literature review

Interactive response systems for educational applications

The IRS adopts specific connecting devices such as Clickers as learning tools to facilitate students’ involvement in class (Pond, 2010). Clickers are a kind of device which can immediately deliver learners’ feedback to instructors, and can help teachers control and understand students’ learning situation. Mayer et al. (2009) adopted a quasi-experimental approach which integrated the use of clickers into a large lecture course in college. The instructors in the clicker group handed out learning sheets, and students used a hand-held remote control device to answer the questions. In the no-clicker group, instructors passed out a sheet containing questions and asked students for a raise of hands for each alternative answer. The findings revealed that the students in the clicker group were more cognitively engaged while learning (Mayer et al., 2009).

Pond (2010) used an interactive response system to improve college students’ performance in an introductory psychology course, and revealed that the students who used the interactive system achieved higher test scores than those who did not. Chuang (2015) developed a mobile-based interactive response system to encourage students to engage in a programming course, and his results pointed out that the mobile-based IRS system triggered collaborative and active learning. Besides, Scrowell (2014) suggested that when using the mobile-based IRS application in the classroom, the class had to have sufficient Wi-Fi access or Internet bandwidth, otherwise the learners might have problems connecting to the IRS server, resulting in negative learning effects.

Overall, the use of IRS applications has the potential to improve classroom learning, and learners tend to have more strongly positive feedback when instructors combine IRS with their teaching or learning strategies (Caldwell, 2007) such as peer learning (Hake, 1998), cooperative learning (Nichol & Boyle, 2003) or the game-based competition learning strategy (Wang, 2015) with IRS applications.

Strategy to foster learning: Game-based strategy with competition

Prensky (2001) stated that teaching with game strategies could address the pitfalls found in traditional education, while Kapp (2012) defined gamification as “using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems.” Games have the power to get learners to learn enthusiastically, and the repetition in gameplay is the driving force that motivates learners to search for target knowledge through the chance to learn by playing (Coyne, 2003). Several studies have explored the pedagogical value of using game elements such as “game-based competition” and a “game-based ranking list” (Huang, 2013; Wei, 2013). For example, Wei (2013) and Huang (2013) pointed out that adopting a competitive mechanism in game activities inspired students’ learning motivation, and the competitive activities has the potential to foster learning and help students concentrate more in class. The factor of beating other competitors in the game is a trigger that can push students to study hard and concentrate on the educational contents (Wang, 2015). Wang (2015) further stated that integrating a game-based element into IRS may have the potential to support flipped classroom learning. Flipped classroom instruction encourages students to study course materials prior to the class, and teachers shift their role from being instructors to information givers (Pardo et al., 2012). Several studies have proposed that the integration of a gameplay element into IRS facilitates the application of flipped classrooms (Chen, Worden, & Bradley, 2015; Davies, Dean, & Ball, 2013) because the game activities and competitive factor have positive effects in terms of triggering students to read the textbook before class in
order to get better game results in the activity. Despite the argument that the integration of a competitive factor into game activities may foster effective learning (Chen & Chen, 2013), there are a few researchers who are concerned that competition in learning may be hurtful or may have negative influences on learners due to their lack of confidence and the social comparison that could result from the learning pressure it creates (Stapel & Koomen, 2005).

Meanwhile, the advantage of the use of a game-based activity is that it helps instructors deliver important topics, and assists learners in remembering the knowledge through the repeated game activities (Wang, 2015). Research had indicated that adopting game-based approach in learning fostered students’ social learning (Tan, Goh, Ang, & Huan, 2013). In social learning theory, students learn by observing others and modify their own behavior accordingly (Bandura, 1977). The game-based activity might be a good learning scenario for learners to involve themselves and learn things through the interaction with peers (De-Marcos, García-Lopez, & Garcia-Cabot, 2016). A number of studies have adopted the interactive response learning system together with game elements in educational scenarios (Méndez, & Slisko, 2013; Wang, 2015). For example, the Quizlet website (Gruenstein et al., 2009) adopts a game element in collecting learners’ instant feedback, and helps teachers to create and employ a variety of online activities in class for fostering learning. Socrative (Méndez, & Slisko, 2013) provides a real-time facility to collect data from students, and offers games to encourage students to answer questions to pass game levels. Van Eck (2006) has stated that learning quality is maximized by leaving the game design of learning up to the teachers, and researchers believe that educational games could be learning aids used as reinforcement to support traditional learning (Tsai, Yu, & Hsiao, 2012) because the games can lower anxiety, enhance involvement, and make learning acquisition more likely, especially for low-achieving students (Young & Wang, 2014). Gamification in educational scenarios helps learners think and learn through game activities, and gamification with competitive activities can motivate learners to be more active participants in the learning process, and thus leads to improvement in knowledge and skills that the learners may not even be aware of (Kiryakova, Angelova, & Yordanova, 2014).

**Advanced learning technology enhances Chinese learning**

Chinese is an important academic subject in several Asian countries, and how to engage students to achieve self-directed learning in lecture classes is an important goal for all Chinese teachers (Chiang, 2014). When teaching Chinese, teachers first provide edification for learners by leading them to understand and appreciate the language. The ability of Chinese article appreciation is achieved by training and enlarging learners’ Chinese characters and vocabulary. After acquiring the form of single characters and vocabulary, teachers gradually lead students to learn the meanings of Chinese contexts, rhetoric and then entire articles (Chi & Chiou, 2015; Chang, 2010). Chinese learning is a multi-phased process (Chi & Chiou, 2015), and researchers have indicated that using advanced educational learning technology can benefit Chinese learning (Chen & Chou, 2007; Chang et al., 2010). For example, Chen and Chou (2007) adopted tablets as learning aids to help students learn Chinese through ubiquitous learning. Hsieh et al. (2010) used the situation learning strategy to help elementary students learn Chinese writing. Chang et al. (2010) developed a Wireless Handheld System to assist senior high school students improve their Chinese reading ability. Their system recorded students’ learning paths and helped Chinese teachers to control the learners’ learning situation. The results indicated that the WHS improved students’, especially novice learners’, Chinese reading ability. Edge et al. (2011) used mobile devices to help learners acquire Chinese characters; Tam and Cheung (2012) employed the i-Write system to support non-native Chinese speaking learners to acquire the correct writing sequence for Chinese characters. Liu, Owen and Sunderraman (2011) developed a flash-based system with a game-based strategy to assist non-native Chinese speakers learning Chinese characters. Sams and Bergmann (2013) developed a note-taking tool that integrated a cooperative learning strategy, the Sharing Unique Reading Feeling system, to help elementary school students acquire Chinese reading skills. These related studies indicated that adopting advanced learning technology fosters Chinese learning; however, not many studies have investigated how learning technology can help with classical Chinese learning.

The above-mentioned studies, which adopted interactive systems to foster learning using questioning in large lecture classes, and gamification and competition factors to assist students’ learning, have individually shown positive effects on fostering students’ learning. Nevertheless, few studies have discussed the cloud-based interactive response system with game-based elements and competitive activities for assisting students’ classical Chinese learning. Thus, the aim of the study was to adopt a cloud-based and cross-device interactive response system in course teaching, and to investigate whether the combined learning strategies could foster students’ language learning, and further, help the students achieve self-directed learning.
Introduction of CCIRS

The system adopted in this study was developed by the team, Kahoot!AS, composed of Johan Brand, Jamie Brooker and Morten Versvik from the Norwegian University of Technology and Science (Figure 1). In this system, instructors can adopt a variety of learning media to create online quizzes (Figure 2). The system supports mobile-based and computer-based web interfaces. The process of using the system starts from the teachers presenting and running the activity projecting questions onto the screen at the front of the classroom. Students can then access the website through personal devices by entering the PIN code (provided by the teacher) and will then see their name appear on the front screen. The system supports competitive and game-based pedagogy. In the activity, the learners play against each other and at the end of each question, students are shown a summary and game-like scoreboard, highlighting the names of the students or groups with the highest scores.

Figure 1. Demonstration of the system interface (from Kahoot!AS https://getkahoot.com/)

Figure 2. Teacher interface for creating a question (from Kahoot!AS http://getkahoot.com/)
Methodology

In this study, comparative test data were adopted to report on the performance of learning classical Chinese in two learning scenarios, with and without CCIRS, and both qualitative and quantitative approaches were employed. The duration of the data collection was four weeks. The learners in the experimental group (E.G.) learned classical Chinese with the assistance of the system, while the control group (C.G.) followed the original classical Chinese teaching method in which the teacher gave lectures and adopted paper-based worksheets to understand students’ learning acquisition. In order to explore how the system facilitated the learners’ classical Chinese learning, the researcher conducted a paper-based Chinese test before and after the experiment as the learning pre-test and post-test. After the experiment, questionnaires were administered to collect the learners’ feedback, and the Chinese instructor was invited to participate in individual interviews to gather qualitative data to support the quantitative information derived from the questionnaire and pre- and post-tests.

Participants

A total of 80 eleventh graders from a senior high school participated in this study and were divided into two groups (E.G. = 40 and C.G. = 40). These two groups of learners had the same Chinese teacher. Learners in both groups were further divided into three subgroups based on their Chinese grades of the previous semester. The grades of the previous semester were the mean values from three Chinese proficiency tests of the last semester. The group of advanced learners were those students whose grades were in the top one third of the class. The grades of the novice learners were in the bottom third of the class, and the rest of the students were categorized as intermediate learners.

The experiment

The duration of the experiment was four weeks. The Chinese instructor asked the learners in the two groups to preview the learning units before the class. Then, during the E.G.’s formal class time, the instructor gave lectures and adopted CCIRS to assist teaching and learning (Figure 4). Using CCIRS, the teacher created time-controlled activities for the students to answer questions. The students answered these questions with their own learning devices. After each activity, CCIRS immediately provided the teacher with a detailed report as an Excel file which included an overview of students’ answers to questions such as how many questions they had answered correctly and their total scores (Figure 3). The detailed report could be viewed as students’ learning portfolios which helped the teacher to know how much the learners in the classroom understood of the learning content so that she could emphasize the content that the learners had failed to grasp. On the other hand, the instructor of the C.G. used paper-based worksheets to test the students’ learning acquisition. The items on the learning sheets for the C.G. were the same as those in the system for the E.G.
The activity on CCIRS in the classroom

The students accessed the given website through their personal devices for entering the activities of CCIRS. There were several learning questions in each activity, and the teacher could embed multimedia contents such as visual combinations and audio objects as question items to present the learning context according to the classical Chinese learning contents (Figure 5). The learners in the activity could acquire the knowledge of classical Chinese literature and the meaning of classical Chinese vocabulary through multimedia materials. During the activity, the learners read the questions on the big screen at the front of the classroom and then chose the right answers to the items. At the end of each question, the teachers were able to view learners’ performance immediately and the system listed the top five winners who answered the question correctly and in the shortest time. The students were motivated to answer questions as quickly as possible in order to move up the on-screen leader board and had their name displayed at the top.
The multimedia contents in the CCIRS

The immediately learning performance of the learners in CCIRS

The top five winners in the activity

Figure 5. The demonstration of the activity (from Kahoot!AS http://getkahoot.com/)

Pre-test, post-test, questionnaire and interview

The pre-test and post-test consisted of multiple-choice questions and short answer questions related to the background knowledge of the targeted Chinese lesson. The items in the pre-test and post-test were validated by the cooperating Chinese teacher. The total score of the test was 100. The perceived learning scales of the questionnaires were modified based on the questionnaire from Carreira’s study for evaluation of students’ language learning motivation and attitudes (Carreira, 2006). The questionnaires in this study consisted of items on a five-point Likert scale and open-ended questions, and the Cronbach’s α of the measures of the questionnaire was 0.91. In order to understand the instructor’s perceptions, opinions and experience of using the system for teaching, the Chinese instructor was invited to take part in an individual focused interview for about an hour (Table 1).

Table 1. Interview questions for the instructor

<table>
<thead>
<tr>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please share your feedback on using CCIRS in your course.</td>
</tr>
<tr>
<td>Please share your opinion on the adoption of the CCIRS system to assist students’ classical Chinese learning, and whether using CCIRS could facilitate Chinese learning? Why and how?</td>
</tr>
<tr>
<td>Please share your findings with us about how the learners’ behavior changed when you used the CCIRS system in your classroom?</td>
</tr>
<tr>
<td>What differences did the use of CCIRS bring to the classroom?</td>
</tr>
<tr>
<td>What are your suggestions about using the CCIRS system for further courses?</td>
</tr>
<tr>
<td>What advantages and disadvantages did the use of CCIRS bring to the classroom?</td>
</tr>
</tbody>
</table>

Data analysis

After deleting the invalid data, the total number of participants was 64 (E.G. = 28 and C.G. = 36). The descriptive statistics, independent t-tests and ANOVA test were adopted for quantitative analysis. For the qualitative data, each participant was given a code. For example, in the code EG-H-01, “EG” represents the
experimental group, “H” stands for advanced learners, and “01” is the student number. The researcher translated the students’ open-ended question feedback to raw data files and re-coded the raw data according to different themes. The final qualitative data were organized and displayed as reduced data from which the findings for each question could be highlighted, as well as for triangulation purposes.

Learning performance of the two groups

The pre-test and post-test were analyzed to answer the first research question. The results of the Levene’s test confirmed that the data met the equality of variance assumption (p = .16 ≥ .05). Both groups showed improvement from the pre-test to the post-test, but the participants did not show significantly better performance in the post-test according to the analysis of independent sample t-tests (Table 2). The ANOVA test with post-hoc comparison was adopted to analyze the students’ performance according to their learning achievement levels. The Table 3 revealed that the factor of various learning achievement levels influenced the students’ learning performance. After doing further factor comparison analysis, it was found that the novice and medium-achievement learners in the E.G. improved significantly in the post-tests, while the advanced learners didn’t showed significant improvement. This indicates that the novice and medium-achievement learners in the E.G. with CCIRS improved significantly in their classical Chinese post-test.

<table>
<thead>
<tr>
<th>Post-test</th>
<th>t-test for equality of means</th>
<th>95% CI of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal variances assumed</td>
<td>1.432</td>
<td>.157</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>1.536</td>
<td>.123</td>
</tr>
</tbody>
</table>

Table 3. Results of the ANOVA tests of learners of various achievements in the two groups

<table>
<thead>
<tr>
<th>Tests of between-subject effects</th>
<th>Mean square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Model</td>
<td>878.980</td>
<td>8.337</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>226920.904</td>
<td>2152.33</td>
<td>.000</td>
</tr>
<tr>
<td>Groups</td>
<td>412.158</td>
<td>3.909</td>
<td>.053</td>
</tr>
<tr>
<td>Learning achievements</td>
<td>1472.122</td>
<td>13.963</td>
<td>.000</td>
</tr>
<tr>
<td>Groups * Learning achievements</td>
<td>375.183</td>
<td>3.559</td>
<td>.035</td>
</tr>
<tr>
<td>Error</td>
<td>105.430</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Multiple comparison

<table>
<thead>
<tr>
<th>Learning achievement and groups</th>
<th>Mean difference</th>
<th>Std. error difference</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced learners</td>
<td>E.G. C.G.</td>
<td>-4.788</td>
<td>4.614</td>
</tr>
<tr>
<td>Medium-achievement learners</td>
<td>E.G. C.G.</td>
<td>10.179*</td>
<td>4.452</td>
</tr>
<tr>
<td>Novice learners</td>
<td>E.G. C.G.</td>
<td>10.082*</td>
<td>4.486</td>
</tr>
</tbody>
</table>

Note. *p < .05.

Students’ perceptions of using CCIRS for classroom learning

The data from the questionnaires regarding learners’ feedback on using CCIRS were analyzed to answer the second question. The overall questionnaire results indicated that the learners were positive about using the system for class interaction (Table 4:Q1) and most of the students reflected that the operation of CCIRS was quite easy (Table 4:Q8). The statistical data from the self-report questionnaires also revealed that the learners liked to participate in the CCIRS activities for classical Chinese learning. Most of the students hoped that the teacher could keep using the system in the following classes (Table 4:Q5&Q18). They highly agreed that the use of the system promoted their interaction with their peers and with the instructor (Table 4:Q14&Q15), and the students expressed their hope that they could use CCIRS for other course subjects (Table 4:Q18). Besides, according to the questionnaire data, the students looked forward to participating in the CCIRS group activities (Table 4:Q17).
I would like to take part in activities in other subject courses. I look forward to the system in the course. I would like to participate in the activity in a group. I would like to participate in the activity personally. The use of the system enhances the interaction between classmates. The atmosphere becomes livelier in class with the use of the system. I would use a nickname to represent myself in the system. I enjoyed using the system in the classical Chinese course. Participating in the activity made me to read more (I will do the preview for the activity). Because the items were too long to read. However, I was more active in activity (EG-L-06). I felt pressure when the system started to countdown during the activity (EG-L-07). Sometimes, my cellophane was crashed and I couldn’t participate in the activity (EG-L-08). Sometimes, the class was out of control (EG-L-09). The classroom was full of noise (EG-L-10). It was a little noisy during the activity (EG-M-07). The schedule of the class was delayed (EG-M-08). The answer items (O, X) in the system were sometime hard to recognize (Only color and figure) (EG-H-04).

**Note.** A.L. = Advanced learner; I.L. = Intermediate learner; N.L. = Novice Learner.

**Table 4. Learners’ feedback on CCIRS**

<table>
<thead>
<tr>
<th>Items</th>
<th>Total</th>
<th>AL</th>
<th>IL</th>
<th>NL</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 I enjoyed using the system in the classical Chinese course</td>
<td>4.50</td>
<td>4.83</td>
<td>4.57</td>
<td>4.50</td>
<td>.758</td>
</tr>
<tr>
<td>Q5 I hope the teacher could keep using the system for Chinese class.</td>
<td>4.23</td>
<td>4.67</td>
<td>4.57</td>
<td>4.11</td>
<td>.501</td>
</tr>
<tr>
<td>Q8 The operation of the system is quite easy.</td>
<td>4.33</td>
<td>4.67</td>
<td>4.36</td>
<td>4.44</td>
<td>.851</td>
</tr>
<tr>
<td>Q12 I would use a nickname to represent myself in the system</td>
<td>4.30</td>
<td>4.33</td>
<td>4.36</td>
<td>4.50</td>
<td>.845</td>
</tr>
<tr>
<td>Q13 The atmosphere becomes livelier in class with the use of the system.</td>
<td>4.60</td>
<td>4.67</td>
<td>4.57</td>
<td>4.72</td>
<td>.731</td>
</tr>
<tr>
<td>Q14 The use of the system enhances the interaction between classmates.</td>
<td>4.47</td>
<td>4.67</td>
<td>4.36</td>
<td>4.67</td>
<td>.672</td>
</tr>
<tr>
<td>Q15 The use of the system enhances the interaction between the teacher and me.</td>
<td>4.10</td>
<td>4.33</td>
<td>4.00</td>
<td>4.44</td>
<td>.486</td>
</tr>
<tr>
<td>Q16 I would like to participate in the activity personally.</td>
<td>3.17</td>
<td>4.17</td>
<td>3.07</td>
<td>3.67</td>
<td>.453</td>
</tr>
<tr>
<td>Q17 I would like to participate in the activity in a group.</td>
<td>4.27</td>
<td>4.33</td>
<td>4.43</td>
<td>4.39</td>
<td>.810</td>
</tr>
<tr>
<td>Q18 I look forward to the Chinese course because of the adoption of the system in the course.</td>
<td>4.07</td>
<td>4.50</td>
<td>4.00</td>
<td>4.33</td>
<td>.648</td>
</tr>
<tr>
<td>Q19 I would like to take part in activities in other subject courses.</td>
<td>4.40</td>
<td>4.67</td>
<td>4.36</td>
<td>4.56</td>
<td>.812</td>
</tr>
</tbody>
</table>

**Note.** AL = Advanced learner; IL = Intermediate learner; NL = Novice Learner.

**Table 5. Learner’s qualitative feedback on CCIRS**

<table>
<thead>
<tr>
<th>Learners’ feedback on the CCIRS (Advantages)</th>
<th>Learners’ feedback on the CCIRS (Disadvantages)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.L. Everybody in the class was crazy about the activity. The course became quite interesting (EG-L-01).</td>
<td>N.L. I felt pressure when the system started to countdown during the activity (EG-L-07).</td>
</tr>
<tr>
<td>I.L. The activity was exciting and it triggered me to preview the learning contents (EG-M-01).</td>
<td>I.L. It was a little noisy during the activity (EG-M-07).</td>
</tr>
<tr>
<td>The system was good and I could remember more important points during the activity (EG-H-02).</td>
<td>The schedule of the class was delayed (EG-M-08).</td>
</tr>
<tr>
<td>It was nervous, exciting and impression. I had interaction with lots of classmates (EG-M-05).</td>
<td></td>
</tr>
<tr>
<td>It made me awaken in the course (EG-M-06).</td>
<td></td>
</tr>
<tr>
<td>A.L. This was my first time to participate in the activity and it was really good (EG-H-01).</td>
<td>A.L. The answer items (O, X) in the system were sometime hard to recognize (Only color and figure) (EG-H-04).</td>
</tr>
<tr>
<td>The system was good and I could remember more important points during the activity (EG-H-02).</td>
<td></td>
</tr>
<tr>
<td>The activity was quite impressed (EG-H-03).</td>
<td></td>
</tr>
</tbody>
</table>

**Learners’ suggestions regarding the system**

It will be better if we could change the answer (within the given time) (EG-L-09).

No suggestion. It is good now (EG-M-10).

*Note.** A.L. = Advanced learner, I.L. = Intermediate learner, N.L. = Novice Learner.
Moreover, according to the open-ended questions, the students reflected that participating in the CCIRS activities in the course was quite exciting and it enhanced their motivation to preview the learning contents. They had good interaction with their peers, and it further promoted their social interaction. The atmosphere in the course was lively, although a few students pointed out that it was noisy and sometimes the class was out of control during the activities (Table 5).

Instructor’s perceptions of adopting CCIRS for teaching and learning

The researcher coded and organized the Chinese instructor’s interview feedback regarding adopting CCIRS in the course activity. The instructor reflected that learners became more confident and concentrated on the course activity, especially the novice learners (Table 6-1). The teacher explained that before the experiment, the novice learners always felt bored or even fell asleep during class; however, they were excited and looked forward to the CCIRS activity every time, and they wanted to share what they had read and studied with each other. Moreover, they became eager to learn or preview the educational contents in order to get good scores on the CCIRS activity (Figure 6-2). The instructor commented that the adoption of CCIRS is a good method for triggering self-study. The atmosphere in the class became energetic, and the learners were very positive about participating in the CCIRS activities. The use of CCIRS therefore clearly has potential for assisting teachers in conducting the flipped classroom strategy.

The instructor also gave suggestions about improving the use of CCIRS such as enhancing the student interface by adding the corresponding name of each answer item so that the system could directly go through all the questions at once without pausing between each question (Table 6-1). Besides, the instructor mentioned that not every class should incorporate CCIRS activities because it takes a great deal of time, and may thus result in learning schedule delay.

Table 6. Summary of the interview feedback from the Chinese instructor

<table>
<thead>
<tr>
<th>1. Feedback regarding using CCIRS in the course</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Using the system in the course is fine, but not in every single class, because when I am in a hurry with the course schedule, the use of the system may take too much time (From interview Question 1)</td>
</tr>
<tr>
<td>• The atmosphere in the class became vivid with the use of the system. Students had a chance to interact with each other, and they clarified the concept through the discussion (From interview Question 2).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Feedback regarding students’ learning performance with the use of the CCIRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>• It is quite a good learning method and the learners were very positive about participating in the activity (From interview Question 2).</td>
</tr>
<tr>
<td>• The low-achievement learners were engaged in the course with the assistance of the system, and before they usually felt bored or even fell asleep in the course; however, they became active in answering the questions in order to get better scores in the activity (From interview Question 3).</td>
</tr>
<tr>
<td>• The students, especially the low achievement ones, became more confident in learning (From interview Question 5).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Suggestions for system modification and instruction applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>• For system modification</td>
</tr>
<tr>
<td>o The current interface of the for teachers is good, and it is easy to input the question items in the system, while the interface for students could be improved by adding the name of each answer item instead of only using colors to represent the answer items (From interview Question 4).</td>
</tr>
<tr>
<td>o Just one suggestion, It would be better if the process of answering questions in the system could directly go through all the questions at one time without pausing between each question. I would like to explain and have discussions about each question at the end of the activity, and I think that would also help the students concentrate on the contents (From interview Question 4).</td>
</tr>
<tr>
<td>• For instruction applications</td>
</tr>
<tr>
<td>o Not every student has a learning device or an Internet connection, and thus it is suggested that the learners be encouraged to participate in activities in groups (From interview Question 1).</td>
</tr>
</tbody>
</table>

Discussion

The use of the cross-device response system facilitates classroom teaching and flexible learning

The use of CCIRS enables instructors and students to overcome the previous hardware and software limitations of IRS. Students using any Internet-connected device with a web browser could participate in the classroom
activities immediately. Before, when an instructor wanted to adopt IRS in the classroom, he/she had to make sure each student had available connecting devices, and the use of the IRS might be limited to a particular brand or mobile operating system. Besides, it took teachers and students extra time to get familiar with the hardware and software tools and to set up the connecting environment prior to the course (Barber et al., 2007). However, the characteristics of the cross-device usage of the CCIRS system enabled students and teachers to access the system more flexibly through various and easily accessible devices. CCIRS reduces the cost burden for students, and the friendly user interface facilitates instructors’ willingness to use the system because there is no need for pre-training in administering the system. Meanwhile, the researcher also noticed that when using mobile devices for IRS course participation, the class has to have sufficient Wi-Fi access, otherwise the learners might have problems connecting to the IRS server, resulting in a negative effect for learning. This finding echoes Stowell’s (2014) previous study findings which indicated the importance of providing learners with sufficient Internet bandwidth to foster IRS activity.

**Game-based competitive activities in CCIRS fosters flipped classroom teaching and learning**

According to the data analysis of the student questionnaires and teacher interview, it was found that the learners were willing to do the preview study in order to achieve good competition results. Such a system can thus be successfully used to motivate students to study the material before class. This finding echoes the results of Huang and Soman (2013) and Kiryakova, Angelova, and Yordanova (2014) which suggested that gamification with competition motivates learners to more actively participate in the learning process. The students reflected that competitive activities with the system enhanced their learning motivation and hence aroused their interest in reading the educational contents prior to the course. The findings of the study also suggest that adopting an interactive response system with the game-based competitive strategy would support flipped classroom teaching. Researchers have indicated that the flipped classroom method facilitates learning by encouraging students to prepare for classes and by providing them with opportunities to gain knowledge before class (Brame, 2013) and in this study, the researcher found that using CCIRS technology with game activities could foster the application of the flipped classroom approach and also promote student-centered learning (Fardoun et al., 2014). The results are also in accordance with Wang’s study (2015) which concluded that interactive response learning has the potential for facilitating flipped learning because the immediate learning activities have positive effects in terms of triggering students to read the textbook before class in order to perform better in the activity.

**Integrating the game-based CCIRS into a classroom promotes novice learners’ learning performance**

According to the qualitative data, the learners in the class were all engaged in the CCIRS course activities. Most of the students liked the ranking list in the system, and mentioned that the atmosphere was livelier in the classroom as a result. Besides, according to the questionnaire results, the findings are in accordance with a previous study which indicated that eagerness to have their names displayed on the scoreboard or ranking list is the main trigger that motivates students to become active learners (Wang, 2015). A previous study also indicated that sometimes noisy and disorganized learning happens when adopting game activities for teaching (Evans, 1979). However, even though we found the same situation that the students were moving around the classroom and the disorganized learning happened during the game-based CCIRS activity, however, the atmosphere in the class was full of joy.

One observation worth noting is the different learning behaviors of the various achievement learners during the course. The novice learners were willing to participate in the learning activities. They gathered around a common screen to discuss the learning content so as to get higher scores in the activity. CCIRS with competitive games fosters students’ social learning and arouses discussion among peers. The instructor indicated that the novice learners were more confident in learning and sharing their thoughts through the game-based competitive activities, and confirmed that the novice learners showed great involvement in the CCIRS activity. Similar results were found in a previous study (Young & Wang, 2014) in which adopting game-based activities motivated less advanced students to engage in the learning activity. However, the findings are contrary to the results of Stapel and Koomen (2005) who indicated that competitive activities may be hurtful to or have negative influences on learners due to their lack of confidence.
Conclusion

In this study, the researcher attempted to answer the questions of whether the adoption of a cross-device interactive response system could improve students’ learning motivation, and whether the integration of CCIRS formed a different classroom teaching and learning pedagogy. To answer these two research questions, the findings of the study demonstrated that the use of CCIRS with game-based competitive activities motivated the learners to acquire knowledge because they felt that they were playing a game instead of taking a test. The students reflected that learning with CCIRS is an entertaining experience, especially for novice and medium-achievement learners who, with the assistance of CCIRS, showed more engagement in the course and achieved significant improvement from the pre- to the post-tests. Van Eck (2006) indicated that “games are effective not because of what they are, but because of what they embody and what learners are doing as they play a game.” The adoption of CCIRS integrated a gameplay element into the course interaction and thus improved learners’ learning outcomes and increased their course engagement. CCIRS is an easy-to-use learning trigger that encourages students to participate in the activities and share their thoughts. Moreover, using CCIRS also arouses course discussion and involvement for the target learning subject and thus enhances students’ social learning. For a long time, previous studies have indicated that the use of interactive response systems has some limitations such as the high cost of devices and setting up the environment. However, the use of this cloud-based IRS is quite intuitive and reduces the loading on the instructor of setting up the application environment. The idea of “bring your own device” (BYOD) could be implemented through the use of CCIRS. Students could use their own mobile phones or tablet PCs to access the activity, while those learners without appropriate devices could use classroom PCs. The instructors just have to log into the web-based interface and create the educational items following their experience of using computers.

In response to the third research question, the CCIRS system integrating the educational strategy, game-based competition, could be an aid used as reinforcement to support flipped classroom learning. The learners were engaged and activated to read and preview the educational contents prior to the course in order to correctly answer the questions raised by the teacher in the class. Besides, teachers could monitor each student’s learning progress from the immediate accountability of responding to questions, and the system could be a tool to help instructors affirm that the learners have absorbed all of the important learning information. Furthermore, the integration of CCIRS might have a positive learning effect on forming flipped classrooms. The use of CCIRS with competitive gameplay in the course may be a suitable trigger for helping learners and instructors conduct flipped classroom applications. The CCIRS pedagogy supports novice learners to participate in learning, and the competitive activities trigger students to become active learners and enhance their social learning.

While the researcher noticed that adopting the game-based elements with CCIRS had the potential for promoting engagement in the classroom, it was also found that the use of the system might delay teaching progress when there is limited time to teach learning contents to students. Consequently, it is suggested that when designing and implementing CCIRS in school courses, the learning schedule of each course must be considered. What matters is whether the advanced tools and learning strategy can be used in interesting ways to promote ways of knowing not possible in existing teaching classrooms.

Limitations and future work

One of the limitations of this study was the small number of students in each subgroup. It is therefore suggested that future studies recruit more participants for confirming the research findings, especially when dividing the learners into subgroups for data analysis. It is also suggested that further studies could focus more on investigating students’ performance and interaction differences with CCIRS. One final limitation of this study is that the target learners and subject of the study were teenagers and classical Chinese; thus, the results cannot be generalized to adult or younger learners or to other learning subjects. Other research issues could include whether students at various educational levels have different perceptions of using CCIRS and how to integrate different teaching pedagogies such as encouraging a learning loop from learners to leaders in the activity on CCIRS. Besides, it is also suggested to investigate whether there are differing learning effects on student’s learning performance and behavior when adopting independent and collaborative learning work with CCIRS using sharable learning tools. Moreover, researchers could explore how CCIRS used together with the flipped classroom approach can foster learning and help students achieve higher order thinking performance, and whether the anonymous use of the system affects students’ learning.
Acknowledgments

The authors would like to thank Ministry of Science and Technology in Taiwan (104-2511-S-032-008 and 105-2511-S-032-002) for the support.

References


Ministry of Education. (2013). 職業學校群科課程綱要之「總綱」 [Syllabus for vocational school]. Retrieved from http://vs.tchvcs.tcu.edu.tw/file/%E8%81%B7%E6%A0%A1%E7%BE%A4%E7%A7%91%E8%AA%B2%E7%8A%8B%E7%B6%B1%E8%A6%81%E7%B8%BD%E7%B6%B1%E6%89%8B%E5%86%8A_1030109%E6%9B%B4%E6%96%B0.pdf


Wei, Y. J. (2013). *Implementing a game-based learning system with a competitive mechanism to enhance students’ learning motivation* (Unpublished master thesis). National TsingHua University, Taiwan.

Creating an Online Learning Community in a Flipped Classroom to Enhance EFL Learners’ Oral Proficiency

Wen-Chi Vivian Wu¹, Jun Scott Chen Hsieh² and Jie Chi Yang²*  
¹Department of Foreign Languages and Literature, Asia University, Taiwan // ²Graduate Institute of Network Learning Technology, National Central University, Taiwan // vivwu123@asia.edu.tw // curtis3883@gmail.com // yang@cl.ncu.edu.tw

*Corresponding author

(Submitted November 5, 2015; Revised June 22, 2016; Accepted August 10, 2016)

ABSTRACT

Since the advent of new technology for learning, innovative language instructors have been constantly seeking new pedagogy to match the potential of technology-enhanced instruction. While previous studies have supported the adoption of technologies to facilitate language teaching and learning, research into enhancing English as a foreign language (EFL) learners’ oral proficiency by creating an online learning community in a flipped classroom remains insufficient. Therefore, the current study examined the impact of an online learning community in a flipped classroom, specifically via mobile platforms, on EFL learners’ oral proficiency and student perceptions. Fifty English-majored sophomores enrolled in two oral training classes at a four-year comprehensive university in central Taiwan participated in this study. A mixed method was employed to analyze multiple sources of data, including pre- and post-tests on oral reading and comprehension questions, a “Community of Inquiry” (Col) questionnaire, and semi-structured focus-group interviews. The results from multiple sources indicated that the online learning community not only facilitated meaningful and positive collaboration but also significantly improved the participants’ oral proficiency, thus leading to more active engagement in highly interactive learning activities, such as storytelling, dialogue collaboration, class discussion, and group presentations.

Keywords
Community of Inquiry, Flipped learning, Online learning community, Oral proficiency

Introduction

The development of communicative competence is an overarching learning objective, and exposure to communicative practice is generally recognized as an essential element of successful foreign language learning and teaching (Council of Europe, 2001). Oral proficiency in a foreign language is the prerequisite for communication of ideas and intelligent conversation. The ability to speak a language is synonymous with knowing the language, since speech is the most basic means of human communication (Folse, 2006). However, inadequate communication and interaction between teachers and students, excessive teacher-led lectures, and relatively fatiguing test-based teaching methods still suppress the development of student communicative competence. Even with years of English learning, English as a foreign language (EFL) speakers still have difficulty mastering English oral skills and are hesitant when speaking English out loud.

Technology, with distinctive features such as mobility, reachability, personalization, spontaneity, and ubiquity, is widely used to facilitate language teaching and learning. In particular, given the benefits and possible learning affordances that mobile-assisted language learning (MALL) offers, incorporating mobile devices appropriately can “have the potential to revolutionize the way we work and learn” (Peters, 2007, p. 1). In recent years, as young users in Asia have been communicating with each other via mobile messaging applications (such as LINE, WhatsApp, and WeChat), research into the role of such instant and text messaging technologies in education has revealed their positive effects on providing platforms for socializing, sharing information, and communicating (Sweeney, 2010).

Flipped learning is an alternative approach that integrates technology into language learning, and that contributes to ample opportunities for students to learn (Chen Hsieh, Wu, & Marek, 2016; Hung, 2015; McLaughlin et al., 2014; Overmyer, 2012). In a conventional class, new knowledge is introduced in the classroom, usually via lecture, and students practice using the knowledge at home, via homework. Flipped learning reverses this paradigm, with information introduced to students before class using technology (such as mobile devices). This allows more advanced learning activities during in-class time, meaning students are given more opportunities to participate in meaningful engaging activities, thus enhancing the learning outcomes (Boucher, Robertson, Wainner & Sanders, 2013).
Studies have shown that flipped learning significantly enhances student learning performance (Chen, Hsieh, Wu, & Marek, 2016; Deslauriers & Wieman, 2011; Hung, 2015; McLaughlin et al., 2014; Sahin, Cavlazoglu, & Zeytuncu, 2015), student engagement (Chen, Hsieh, Wu, & Marek, 2016; Jamaludin & Osman, 2014), and produces enhanced learning outcomes (Chen, Hsieh, Wu, & Marek, 2016; Baepler, Walker & Driessen, 2014; Moravec, Williams, Aguilar-Roca, & O'Dowd, 2010). Bishop and Verleger (2013) contended that a flipped classroom is an educational technique that consists of two important components: (1) the use of computer technologies such as video lectures, and (2) the involvement of interactive learning activities. In fact, flipped learning effectively cultivates student autonomy and arouses student awareness (Yang, 2013), by allowing students to “proceed at their own pace, guide themselves to additional content, and assess their own learning gains” (McLaughlin et al., 2013, p. 196). Furthermore, flipped instruction provides autonomous supportive learning contexts that not only adopt students’ different perspective and thoughts but support students’ autonomous self-regulation (Reeve, 2009). Simply put, the core of flipped learning is to provide a learning community where students develop knowledge through constructive learning experiences, peer interaction and, collaboration.

Chen, Wang, Kinshuk, and Chen (2014) expanded ideas about flipped learning with their “seven pillars of flipped learning” model (Figure 1). The instructional design of the current study, in which LINE (a mobile application that features instant communications among electronic devices such as smartphones, tablet computers, and personal computers) was used to form an online learning community for flipped learning, fits well with the FLIPPED schema. Both an online learning community and a physical classroom gave students a flexible environment. A student-centered learning culture allowed students to participate actively in collaborative and interactive instructional activities outside the classroom by means of the online learning community. Intentional content made by instructors in the form of videos was specifically designed for students to learn and explore new knowledge. The researchers were all professional educators who observed and monitored student engagement and progress, offered online and in-person feedback, and evaluated the students’ performance. The learning activities provided progressive networking learning activities in which the students acquired knowledge, interacted, and collaborated with their partners via the online learning community. The professional educators in the current study not only possessed good instructional skills, strategies, and attitudes prerequisite for a positive learning environment, but also were aware of the transactional distance by considering the proper combinations of structure, dialogue, and learner autonomy, providing engaging and effective learning activities. Finally, the ubiquitous online learning community via LINE combined with advanced in-class learning activities provided a diversified and seamless learning platform.

An interactive online learning community allows students to develop strong relationships with fellow students (Murdock & Williams, 2011). It provides learners with opportunities to meet regularly with their partners for collaborative construction and improvement of knowledge about chosen topics. A great deal of literature suggests that online learning communities are an effective way to promote the sharing and building of knowledge by learners (Ke & Hoadley, 2009), to enhance overall learning and critical thinking, to foster active learning, and to develop more positive learning attitudes, in comparison with conventional classes (Gazi, 2009). The Community of Inquiry (CoI) framework is a widely-used model for examining and evaluating a learning community (Garrison, Anderson, & Archer, 2001; Garrison, Cleveland-Innes, & Fung, 2010). CoI focuses on the intentional development of the community with an emphasis on the processes of instructional conversations that lead to epistemic engagement (Shea & Bidjerano, 2010). CoI encompasses three interdependent elements that facilitate meaningful online learning: teaching presence, social presence, and cognitive presence. The teaching presence, referring to how instructors sequence the learning activities and facilitate learning (Koh, Herrring, & Hew, 2010), encompasses the design, direction, and support of student activities to provide a powerful learning
experience (Rubin, Fernandes, & Avgerinou, 2013). *Social presence*, the most effective way to support the social and interpersonal communication required for online teaching and learning (Lowethal & Dunlap, 2010), includes affective responses and expression, open communication with others during the course, and cohesive communicative responses (Díaz, Swan, Ice, & Kupczynski, 2010; Rubin, Fernandes, & Avgerinou, 2013). *Cognitive presence* refers to the development of critical thinking skills (Scherer Bassani, 2011); the engagement with course concepts; and the ability to create meaning out of ideas, develop and build competence via discussion, and reflect and apply the newfound meaning (Rubin et al., 2013). Such presence is evidenced in a collaborative constructivist learning environment or in online discourse where students share related experiences via critical thinking to achieve a shared understanding (Burgess, Slate, Rojas-LeBouef, & LaPrairie, 2010; Garrison et al., 2001; Yang, Quadir, Chen, & Miao, 2016).

In recent years, communication via mobile messaging applications (such as LINE, WeChat, and WhatsApp) has gained increasing popularity among young users in Asia. Previous research into the effects of instant and text messaging technologies in education has shown that such applications serve as online media for socializing, information sharing, and communicating (Sweeny, 2010), and result in stronger motivation and support (Chen Hsieh, Wu, & Marek, 2016; Coniam & Wong, 2004; Wu, Marek, & Chen, 2013). However, there has been little investigation into using MALL for online learning communities. Investigation into flipped learning is also relatively new (Ash, 2012; Bergmann & Sams, 2012; Herreid & Schiller, 2013; Tucker, 2012), let alone an in-depth probe into whether the integration of an online learning community into flipped learning via mobile platforms can enhance the oral proficiency of EFL learners. In addition, research on use of the voice function that many social media platforms provide remains scarce. To scrutinize the benefits and affordances of online learning communities, mobile technologies, and flipped learning, the current study examined the effect of using an online learning community for flipped learning, specifically via the smartphone app LINE, on the oral proficiency and learning perceptions of EFL learners.

Accordingly, the researchers employed a mixed-method research design to study how an online learning community in a flipped classroom incorporating online verbal interaction, measured via CoI, impacted the EFL learners’ perceptions and oral proficiency, all compared to conventional learning. The following research questions guided the study:

- Were there any differences in the participants’ oral proficiency between the two instructional methods (flipped and conventional learning)?
- Were there any differences in the teaching/social/cognitive presences between the two instructional methods?
- What were the participants’ overall experiences learning English via flipped learning?

This study is significant and at the cutting edge because different from previous studies, it probed into all of the three presences (teaching, social, and cognitive) in the CoI framework that an innovative flipped instruction created, specifically examining the use of an online learning community via the smartphone app LINE in a flipped classroom in EFL oral training classes at the university level. Moreover, the current study provides a holistic flipped instructional design that integrated the four skills of English as a whole, in which passive learning activities such as unidirectional lectures were replaced by instructional videos and collaborative activities before class, allowing precious class time to be spent on interactive and collaborative learning activities.

**Methods**

**Participants**

The participants in the current study were 50 English-major sophomores enrolled in required English Oral Training classes at a four-year university in central Taiwan, mostly female and between the ages of 20 and 21. The participants had studied English for around eight years through high school English education, with nearly 50% of them passing the Intermediate Level of the General English Proficiency Test (GEPT) and some passing the first stage of the High-Intermediate Level. Their English proficiency level was considered to be upper-intermediate, indicating a capacity to make inquiries and hold conversations on daily topics, describe and express topic-specific opinions, and share personal thoughts and viewpoints in social interactions without much difficulty. The participants experienced conventional lecture-based instruction for the first eight weeks of the study and then shifted to flipped instruction for an additional eight weeks.
The researchers chose this model to align with the Within-Subjects research methodology (Creswell, 2013) so that the students would all have experience with both instructional styles. As such, the study was not a formal experimental design and it was not a research design examining the effects of variables. It was not necessary, therefore, to have separate control and experimental groups.

**Instructional design**

To fit the nature of this oral training class, the instructor chose the GOOD CHATS (3rd ed.) textbook, an English conversation textbook designed for advanced students of English, because it covers the most modern and frequently used English idioms, expressions, and turns of phrase useful for oral communication. GOOD CHATS, with its goal set for students to communicate their own ideas or opinions in English, includes 15 topics and each topic features (1) reading passages with idioms, phrases, and collocations useful for the unit topic, and (2) participatory guided dialogues. Six units were chosen (Work, Sports, and Money for the conventional instruction; Time, Childhood, and the Internet for the flipped instruction), each of which included a reading passage, four comprehension questions, 30-40 relevant idioms, and a guided dialogue (Chat for Two) that required the students to draft the dialogue collaboratively.

---

**Procedures of the flipped instruction**

1. **Before class**
   - Students preview video clip on idiom introduction and post personal short stories (multiple trials of audio recording) on LINE groups
   - Students provide verbal feedback to partners on LINE
   - Instructor gives feedback on LINE

2. **First 100-minute class meeting**
   - Misconception clarification
   - Students’ short story presentation
   - Collaborative and engaging learning activities (Problem-based discussion & group presentation)

3. **Before class**
   - Students preview video clip on reading texts
   - Students post guided dialogues (multiple trials of audio recording) on LINE
   - Instructor gives feedback on LINE

4. **Second 100-minute class meeting**
   - Instructor goes through the reading text
   - Instructor engages students in discussion of comprehension questions
   - Students express their opinions, either individually or in groups
   - Students’ oral presentation of guided dialogues

---

*Figure 2. Procedure of the current study*
The LINE smartphone app was chosen to form online learning communities because it appeared to provide the affordances required for the instructional design of this study. Line is cross-platform (Android, iOS, and PCs) and has high popularity among over 600 million users (Eun-ji, 2015). It features text messaging between individuals and among groups as well as sending audio messages and files. The students in this study were highly familiar with the app and thus required no extra training in its functions. In the flipped instruction, the participants were randomly paired and established two-person LINE groups. Having previewed the instructional videos on useful idioms and reading texts covered in those chosen units, the participants completed their personal story writing in English and collaborated with their partners in English over guided dialogues. They then recorded their stories and guided dialogues with the voice input function provided by LINE and uploaded their English audio messages to their LINE groups. To enhance their English oral proficiency, they were encouraged to have multiple trials of such audio recording until they found it satisfactory, during which time their partners gave feedback in English accordingly. The researchers also joined the participants’ LINE groups to give feedback, to observe online peer interaction, and to monitor the overall progress. Figure 2 displays the instructional procedure and data collection process of the current study.

Since the participants had already studied the reading passages and relevant idioms, the instruction in the classroom moved from detailed lecture about micro-level grammar rules to an interactive discussion over meaning clarification and the content of the chosen topic. All instruction was conducted in English. In-class English-based collaborative learning activities included discussion and group presentation, in which the participants were given ample opportunities to apply what they learned in meaningful and authentic settings. After finishing each set of three topics, the participants completed a post-test, identical to the pre-test, which was recorded and graded against the chosen scoring rubric.

During the conventional instruction units (also conducted in English), on the other hand, the instructor elaborated on the reading contents for each chapter during class meetings. Although the participants also experienced collaborative activities like those in the flipped instruction, the need to explain the idioms and introduce the reading passage in class left considerably less time overall for higher-level learning activities.

Research design

Multiple sources of data collection examined the participants’ perceptions of the flipped learning experience, including (1) pre- and post-tests of oral reading and comprehension questions, (2) the CoI survey, (3) two semi-structured focus-group interviews, and (4) in-class observations by the instructors. Figure 3 illustrates the interplay among the issues explored, research questions, and data collection.

Quantitative data analysis

To examine the participants’ overall oral fluency, in answer to research question one, the participants completed pre- and post-tests of oral reading and comprehension questions selected from the textbook. The respective pre-
tests and post-tests, for the conventional instruction and the flipped instruction (see Appendix A), were identical in content. The participants responded orally by reading one paragraph aloud and answering one comprehension question for each of the three units. The classroom instructor and the instructor’s research assistant evaluated the pre-test and post-test audio recordings. To assure higher inter-rater reliability evaluated via Krippendorff’s alpha, the researchers used the IELTS Assessment Criteria: Speaking to evaluate the participants’ oral performance, covering (1) fluency and coherence, (2) lexical resource, (3) grammatical range and accuracy, and (4) pronunciation. Inter-rater reliability, measured with Krippendorff’s alpha at .80, suggested a good reliability (Hayes & Krippendorff, 2007). Since all the participants were upper-intermediate and could read as well as respond to the requirements of the tests, the researchers removed the lowest three point levels or “bands” of the IELTS rubric, leaving seven point categories in the rubric. For the oral reading section, two criteria were employed, including fluency/coherence, and pronunciation. Although three reading passages were chosen from different topics, the participants’ oral reading was evaluated as a whole since the participants’ learning task was to read aloud. Therefore, the overall score of the reading section accounted for 14 points. Unlike the reading section that examined the participants’ oral reading ability as a whole, the four IELTS assessment criteria were applied to each of the three comprehension questions since the participants’ responses to each question were topic-specific. With each question accounting for 28 points, the overall score of this part was 84. Therefore, the total score for the oral test was 98. The means of the pre- and post-tests were calculated to compare differences (i.e., flipped versus conventional). Furthermore, a Paired-Samples t-Test was employed to investigate the participants’ oral learning outcomes in two different forms of instruction.

In answer to research question two, the CoI Scale (Arbaugh et al., 2008) in the form of a 5-point Likert scale (see Appendix B) was adopted to investigate the participants’ perceptions about the differences between the conventional lecture-based instruction and the flipped experience. The overall reliability of the CoI scale was greater than .90, and the Cronbach’s alpha values for the teaching, social, and cognitive presences were .94, .91, and .95, respectively, suggesting high internal consistency of the CoI scale. Descriptive statistics were used to examine the participants’ responses to the three elements: teaching presence (items 1-13), social presence (items 14-22), and cognitive presence (items 23-34).

Qualitative data analysis

Two semi-structured focus-group interviews with protocols developed by the researchers were adopted to answer research question three regarding the participants’ overall perceptions of the flipped instruction in the course. The interviews focused on the perceptions of (1) the flipped instructional design as a whole (e.g., improvement in oral ability, boring, interesting, challenging, easy…), (2) watching the video clips (learning materials) ahead of class time as well as the subsequent exercises (e.g., story writing and guided dialogue practice on LINE), and (3) the interaction with, and feedback from partners and the instructors. In addition, the in-class interactions among the participants were also recorded and analyzed to showcase the participants’ learning experiences.

Results and discussion

The overall analysis of the comparison between the pre- and post-tests, the CoI survey, the focus-group interviews, and the instructor’s in-class observation indicated that the use of LINE as an online learning community for flipped instruction created a self-paced learning context. It also created a collaborative constructivist learning environment, in which meaningful learning tasks prior to the class and interactive learning activities in class enhanced the participants’ oral fluency and accuracy and made students more engaged in learning (both at home and in class). The results of the current paper are presented in accordance with the research questions.

RQ1: Were there any differences in the participants’ oral proficiency between the two instructional methods (flipped and conventional learning)?

Descriptive statistics comparing the pre- and the post-tests in the conventional and flipped instruction units indicated that in both types of instruction, the mean score of the post-test was higher than that of the pre-test (see Table 1). The mean score of the flipped learning ($M = 85.98$) was much higher than that of the conventional instruction ($M = 66.6$). While the maximum scores of the pre-tests in both categories of instruction did not differ greatly, the maximum score of the post-tests in the flipped instruction unit ($M = 95$) was considerably higher than that in the conventional instruction unit ($M = 82$).
Table 1. Descriptive statistics of the pre-test and the post-test

<table>
<thead>
<tr>
<th>Test</th>
<th>Instruction</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>Flipped</td>
<td>50</td>
<td>69.94</td>
<td>5.80</td>
<td>59</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>50</td>
<td>59.29</td>
<td>5.98</td>
<td>48</td>
<td>74</td>
</tr>
<tr>
<td>Post-test</td>
<td>Flipped</td>
<td>50</td>
<td>85.98</td>
<td>5.58</td>
<td>75</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Conventional</td>
<td>50</td>
<td>66.6</td>
<td>5.92</td>
<td>55</td>
<td>82</td>
</tr>
</tbody>
</table>

The Paired-Samples t-Test shown in Table 2 indicated that in both forms of instruction, the participants performed significantly better on the post-test ($p < .001$) compared to the pre-test, and that the post-test of the flipped instruction was higher than that of the conventional instruction at a significant level ($p < .001$). These results indicated that while both methods of instruction were effective in enhancing the participants’ oral proficiency, the flipped instruction contributed to significantly better learning outcomes than the conventional lecture-based instruction ($p < .001$), echoing the findings of previous studies that have shown the positive effect of flipped learning on learning outcomes (Chen Hsieh, Wu, & Marek, 2016; Fulton, 2012; Sahin, Cavlazoglu, & Zeytuncu, 2015; Strayer, 2012; Zappe, Leicht, Mssner, Litzinger, & Lee, 2009). Compared with the mean differences in the conventional instruction, the wider range of mean scores between the pre- and post-tests in the flipped instruction suggested that the participants learned more from the flipped instructional design, because of the ample opportunities for conversational applications in authentic, supportive, interactive, engaging, and collaborative learning contexts.

Table 2. Paired-samples t-test of the pre-test and the post-test

<table>
<thead>
<tr>
<th>Paired differences</th>
<th>Mean</th>
<th>SD</th>
<th>Std. error mean</th>
<th>95% CI of the difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flipped: Post to Pre</td>
<td>16.04</td>
<td>1.64</td>
<td>.24</td>
<td>15.57</td>
<td>16.52</td>
<td>67.901***</td>
<td>49</td>
</tr>
<tr>
<td>Conventional: Post to Pre</td>
<td>7.31</td>
<td>1.34</td>
<td>.20</td>
<td>6.92</td>
<td>7.70</td>
<td>37.826***</td>
<td>49</td>
</tr>
<tr>
<td>Post (flipped) to Post (conventional)</td>
<td>19.38</td>
<td>1.32</td>
<td>.20</td>
<td>18.99</td>
<td>19.76</td>
<td>102.094***</td>
<td>49</td>
</tr>
</tbody>
</table>

Note. ***$p < .00$. 

RQ2: Were there any differences in the teaching/social/cognitive presences between the two instructional methods?

The CoI examined whether the flipped instruction online learning community made any differences in the teaching, social, and cognitive presences by asking the participants to compare the learning experience in the conventional lecture-based instructional unit with that in the flipped instruction. To begin with, the results in Table 3 revealed that the average scores of the three presences in the post-survey were higher than those in the pre-survey. The social presence held the highest mean difference among the three presences, highlighting beneficial open communication and interpersonal relationships created by the online learning community in the current flipped instruction.

Table 3. Mean differences of the three presences

<table>
<thead>
<tr>
<th>Survey</th>
<th>Teaching</th>
<th>Social</th>
<th>Cognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-</td>
<td>3.53</td>
<td>3.47</td>
<td>3.42</td>
</tr>
<tr>
<td>Post-</td>
<td>4.05</td>
<td>4.11</td>
<td>4.04</td>
</tr>
<tr>
<td>Post-pre</td>
<td>0.52</td>
<td>0.63</td>
<td>0.62</td>
</tr>
</tbody>
</table>

The Paired-Samples t-test shown in Table 4 indicated that significant differences in the three presences between the flipped instruction and the conventional lecture-based instruction, suggesting that the participants had a significantly more positive perception of the flipped learning compared to conventional learning.

In the teaching presence (see Table 5), the average scores of all 13 items in the post-survey were higher than those in the pre-survey. Item 12, “In the previous/current instruction, the instructor provided feedback that helped me understand my strengths and weaknesses relative to the course’s goals and objectives,” had the highest mean difference. This indicated that the participants recognized the usefulness of the instructor’s timely online feedback as well as face-to-face comments in guiding them to become more metacognitively aware of their capabilities and to locate appropriate responding strategies. Furthermore, the responses to Item 7 and Item 9
revealed that they became more engaged in productive dialogue and to explore new concepts in the flipped instruction, because their pre-class efforts in learning idioms and previewing reading contents, in turn, generated more time for interactive learning activities.

Table 4. Paired-samples t-test of the pre-survey and the post-survey of CoI

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Std. error</th>
<th>95% CI of the difference</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post to Pre (Teaching)</td>
<td>.52</td>
<td>.55</td>
<td>.08</td>
<td>.35</td>
<td>.69</td>
<td>6.106***</td>
<td>49</td>
</tr>
<tr>
<td>Post to Pre (Social)</td>
<td>.63</td>
<td>.55</td>
<td>.10</td>
<td>.46</td>
<td>.81</td>
<td>7.516***</td>
<td>49</td>
</tr>
<tr>
<td>Post to Pre (Cognitive)</td>
<td>.62</td>
<td>.60</td>
<td>.09</td>
<td>.43</td>
<td>.81</td>
<td>6.636***</td>
<td>49</td>
</tr>
</tbody>
</table>

Note. ***p < .001.

Table 5. Mean differences of the items in the teaching presence

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.71</td>
</tr>
<tr>
<td>Post-</td>
<td>4.1</td>
<td>4.05</td>
<td>4.24</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.17</td>
<td>4.1</td>
<td>4.07</td>
<td>3.98</td>
<td>4</td>
<td>4</td>
<td>3.86</td>
</tr>
<tr>
<td>Post-pre</td>
<td>0.38</td>
<td>0.48</td>
<td>0.52</td>
<td>0.12</td>
<td>0.57</td>
<td>0.38</td>
<td>0.64</td>
<td>0.5</td>
<td>0.71</td>
<td>0.5</td>
<td>0.57</td>
<td>0.81</td>
<td>0.55</td>
</tr>
</tbody>
</table>

The results of the social presence shown in Table 6 resembled those of the teaching presence in that the average scores of all 9 items in the post-survey were higher than those in the pre-survey. Item 18, “I felt comfortable participating in the course discussions,” had the highest mean difference, pointing out the supportive interactive context created in the flipped instruction design. The participants’ responses to Item 17 and 19 also suggested that the flipped instruction generated a more comfortable conversation and interaction with their peers, in comparison with the conventional lecture-based instruction.

Table 6. Mean differences of the items in the social presence

<table>
<thead>
<tr>
<th>Item</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-</td>
<td>3.5</td>
<td>3.64</td>
<td>3.76</td>
<td>3.24</td>
<td>3.17</td>
<td>3.36</td>
<td>3.26</td>
<td>3.4</td>
<td>3.93</td>
</tr>
<tr>
<td>Post-</td>
<td>4.05</td>
<td>4</td>
<td>4.17</td>
<td>4.17</td>
<td>4.38</td>
<td>4.17</td>
<td>3.93</td>
<td>3.95</td>
<td>4.17</td>
</tr>
<tr>
<td>Post-pre</td>
<td>0.55</td>
<td>0.36</td>
<td>0.4</td>
<td>0.93</td>
<td>1.21</td>
<td>0.81</td>
<td>0.67</td>
<td>0.55</td>
<td>0.24</td>
</tr>
</tbody>
</table>

With respect to the cognitive presence demonstrated in Table 7, the average scores of all 12 items in the post-survey were higher than those in the pre-survey, among which Item 34, “I can apply the knowledge created in the current class to my work or other non-class related activities”, had the highest mean difference. The results indicated that the participants were able to use what they had learned in wider applications, a noteworthy point in the field of language education because the true goal of language learning is daily communication rather than for passing tests or certificates. The participants also indicated that they felt motivated to explore content related questions. What’s more, the flipped instruction provided the participants with opportunities to synthesize and reflect on new information, leading to a better understanding of the fundamental concepts of the course.

Table 7. Mean differences of the items in the cognitive presence

<table>
<thead>
<tr>
<th>Item</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-</td>
<td>3.5</td>
<td>3.52</td>
<td>3.48</td>
<td>3.52</td>
<td>3.38</td>
<td>3.48</td>
<td>3.43</td>
<td>3.48</td>
<td>3.33</td>
<td>3.36</td>
<td>3.45</td>
<td>3.17</td>
</tr>
<tr>
<td>Post-</td>
<td>4.12</td>
<td>4.05</td>
<td>4.21</td>
<td>4.02</td>
<td>4</td>
<td>3.98</td>
<td>4.14</td>
<td>4.05</td>
<td>4.05</td>
<td>3.93</td>
<td>4</td>
<td>3.95</td>
</tr>
<tr>
<td>Post-pre</td>
<td>0.62</td>
<td>0.52</td>
<td>0.73</td>
<td>0.5</td>
<td>0.62</td>
<td>0.5</td>
<td>0.71</td>
<td>0.57</td>
<td>0.71</td>
<td>0.57</td>
<td>0.55</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Evaluation of the teaching presence included aspects such as instructional design and organization of learning activities, discourse facilitation, and direct instruction via the online learning environment, revealing that the participants benefited from explicit curriculum and interactive learning activities, and from ample instructor-created opportunities for collaboration and reflection on learning. The findings were consistent with previous research in that teaching presence is a core factor in determining learning outcomes, because students benefited from the instructor’s provision of clear guidance and direction in the technology-enhanced instruction. Such presence assisted learners in accomplishing personally meaningful and educationally worthwhile learning outcomes (Anderson, Rourke, Garrison, & Archer, 2001).
The social presence, which is defined as the degree to which participants feel affectively connected to one another and examined via open communication and interpersonal relationships, supported the social and interpersonal communication required for online learning. The results of the current study echo the findings of previous studies that learners are engaged in a continuous process of maintaining interpersonal relationships, identifying with the community, and involving in meaningful and trustworthy communication (Garrison & Akyol, 2013). With respect to the cognitive presence, evaluated via the development of critical thinking skills, meaning creation out of discussion, and meaning application, the current study yielded a positive learning experience that guides learners through a triggering event, exploration, integration, and resolution phases (Garrison & Arbaugh, 2007). The two presences combined such that the participants shared related experiences to reach a shared understanding via critical thinking. The results align with the findings by Garrison et al. (2010), suggesting that online learning communities can employ the social presence, serving as an important factor facilitating student learning with technology-enhanced instruction. Furthermore, the social presence serves as an important prerequisite for collaboration and critical discourse (Garrison, 2011), helping the participants in the current study to project their personal characteristics into the learning community through the use of affective expression, open communication, and various means to establish group cohesion (Garrison, Anderson, & Archer, 1999; Garrison & Arbaugh, 2007; Rubin et al., 2013). The participants gradually gained an increased understanding and awareness of the learning materials via the cognitive presence, which in turn enabled them to opt for learning strategies appropriate to the given learning tasks (Garrison & Akyol, 2015).

**RQ3: What were the participants’ overall experiences learning English via flipped learning?**

Based on the analyses of the semi-structured focus-group interviews that explored the participants’ reflections about the flipped instructional design as a whole, instructional videos and subsequent activities employed in the study, and interaction as well as feedback involved in the current study, four closely-related recurring themes were identified, including (1) time engagement in the flipped instruction, (2) effectiveness of the flipped instruction on learning outcomes, (3) lively interaction and constructive feedback in the online learning community, and (4) autonomous learning.

**Time engagement in the flipped instruction**

In comparison with the conventional lecture-based instruction in which students had insufficient opportunities to be engaged in pre-lesson activities, instructors needed to lecture for much of the class time, and students had a limited time span for activity-based learning, most participants reported far more time engagement in the flipped instructional design. One student noted, “I spent more time than before. I watched the videos on YouTube and I also discussed with my partner before we did the work on LINE.” Another student said, “Compared with the previous conventional design, I actually spent more time in the flipped teaching. I tried to know the reading content, so that I could speak more in class.” Still another responded that, “I had to spend much more time than usual. But, that is why I can learn much more effectively.”

**Effectiveness of the flipped instruction on learning outcomes**

Most students thought that, compared with the conventional lecture-based instruction which started normally with unidirectional and didactic lectures, the flipped instruction more effectively and efficiently enhanced their oral performance. One student noted that, “Such learning released my tension. After learning the information covered in the instruction, I felt more confident when communicating with my peers in English. I can listen to my own recording and take my partner’s recording for reference. I can practice as much as I want.” Another expressed personal thoughts by saying that, “Through this class, I’m not afraid of speaking in English. I got many chances to speak, write, and read in English. I didn’t know many English-related expressions before, so my English expression was simple. But now I can use a wide variety of expressions.” Some students mentioned the ample practice they had to enhance their oral proficiency by stating that, “Recording through LINE helps me a lot, since I can record as many times as I wish, until I am satisfied with my recording.” Some students recognized the effectiveness of the flipped instruction by noting that, “I can learn English through LINE anytime and anywhere,” and “I have learned much more from the teacher, the instructional videos and the interaction with my partner. And I become more active in learning English.” One student concluded that, “Sometimes I couldn’t get the meanings by reading alone. But the instructional videos helped me understand the overall meaning, with which I could apply what I’ve learned to my story writing, dialogue practice, and daily conversation.”
Lively interaction and constructive feedback in the online learning community

In the interview, students mentioned the interaction with their partners in the pre-class tasks, in-class activities, and assignment not only led them to be more engaged but also provided more learning opportunities to use English in a more authentic way, as illustrated by one student’s response that “I particularly loved the interaction and recorded the guided dialogue with my partner on LINE, because it provided a natural context for both of us to learning English. It’s all about communication and interaction.” Another commented that, “Although I spent a lot of time in this course, I have benefited from interacting with my partner and the instructor. I could practice those useful expressions with my partner and in class.” In addition, the feedback from the instructors and their partners were considered beneficial because such suggestions corrected their inaccurate usages. As one student put it, “I could immediately discuss with my partner about inappropriate expressions and revised my work accordingly.”

Autonomous learning

More time engagement in the instruction, lively interaction, and constructive feedback gradually guided the students to become autonomous learners. “I prefer such flipped learning experience since I could get immediate feedback and assistance, which enhances autonomous learning and absorption of knowledge,” commented one student. Some students noted that the partnership actually drove them to be more responsible for learning, since they had to keep reminding themselves that their partners were also learning along with them, as reflected in one student’s response that, “If I am not well-prepared, I might become a drag on mutual learning.”

The online learning community, embodied in the authentic interaction and communication via LINE, resembles the findings of previous studies in that the students showed positive acceptance of the flipped instruction (Fulton, 2012; Lucke, 2014; Mortensen & Nicholson, 2014; Murdock & Williams, 2011), because valuable instructional time was mostly spent for meaning clarification, problem solving, and interactive collaboration (Bishop & Verleger, 2013; Boucher et al., 2013; Cole & Kritzer, 2009; Simkins & Maier, 2010). The instructional videos viewed outside the classroom served as mechanism to “fill the vacuum” that instructors would otherwise have left (Tucker, 2012, p. 83), freeing class time for more advanced learning and making the students more engaged in meaningful, interactive, and collaborative activities for deeper concept learning. Pierce and Fox (2012, p. 4) commented that “quality, not necessarily the quantity, of student-teacher interaction is a compelling force in improving student performance.” In the flipped instruction employed in the current study, the instructor served more as a guide than an authority, guiding students to build confidence and learn actively (Sarawagi, 2014). In addition, through critiquing the partner’s works and providing feedback, students grew as they moved from micro-level issues such as focusing on grammar to macro-level aspects such as idea consolidation, echoing Findlay-Thompson and Mombbourquette’s statement that “application, analysis, and evaluation” is the outcome cultivated in a flipped course redesign (2014, p. 65). Since students were encouraged to take responsibility for their learning outside the classroom and to participate actively in the instructional activities, they were gradually cultivated to be self-directed in, and responsible for, their learning (Boucher et al., 2013; Overmyer, 2012), thus leading to the enhancement of autonomous learning.

Conclusions

The results of the current study revealed that the participants’ oral proficiency was significantly enhanced as the result of the employment of an online learning community via LINE in the flipped instructional design. Constructive, collaborative, contextual, and self-directed pre-course tasks as well as in-class activities effectively motivated the participants to be more engaged and encouraged them to apply what they have learned to authentic settings. The participants also expressed their positive perception of the flipped instruction adopted in the current study, because such instruction made significant differences in the teaching, social, and cognitive presence while the participants made a comparison between the current flipped instruction and conventional lecture-based teaching. Overall, the online learning community in the flipped instruction not only led to meaningful learning while facilitating positive interaction and collaboration, but also significantly enhanced the participants’ oral proficiency, making them more competent in learning activities, such as storytelling, dialogue interaction, class discussion, and group presentations.

Based on the findings and subsequent discussion of the current study, the researchers offer the following recommendations for practice.
Flipped learning could be effectively applied to language teaching and learning in an EFL context. Such instructional design enabled students to be motivated and engaged in learning activities, thus facilitating English teaching and enhancing learning outcomes.

The mobile-assisted online learning community with the use of LINE adopted in this study is an appropriate instructional design, as it provides an authentic setting for genuine interaction among students. The teaching presence, social presence, and cognitive presence could be significantly enhanced as students are engaged in language exchanges with their peers in the online learning community.

Students’ responsibility for performing the flipped learning activities outside of class is essential. To make sure students complete required learning tasks prior to instructional meetings, teachers must have ways of monitoring students’ progress throughout the whole process.

It is hoped that the results of online learning community yielded in the current study open the way for further research and for integration of innovative instructional designs in an EFL setting. Future studies might focus on the extent to which learners of different proficiency levels benefit from an online learning community or a flipped instruction, or scrutinize the dynamic interaction in an online learning community.

References


Gazi, Z. A. (2009). Implementing constructivist approach into online course designs in Distance Education Institute at Eastern Mediterranean University. The Turkish Online Journal of Educational Technology, 8(2), 68-81.


Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. Educational Psychologist, 44(3), 159-175.


Appendix A

The pre-test / post-test for the conventional instruction

I. Paragraph reading:

1. After one finds a position, achieving job satisfaction depends a lot on being able to cope with the demands of the job. Sometimes a lack of necessary interpersonal skills can create problems. Some people may be highly productive, but if they are unable to work well with other people, they may find it difficult to get ahead. They may not be suited to having managerial positions. For a manager, the inability to delegate responsibility to others makes it necessary to work overtime; by taking on too many responsibilities, the manager may create a stressful job situation for everyone. Rather than advance in the company, this “workaholic” may experience job burnout within a few years and have to seek a less demanding job to alleviate the stress, if he or she doesn’t have a heart attack first.

2. Not only individual schools, but states and nations may differ in the amount of emphasis they place on athletic development. Some governments provide funding for athletic programs, often with an eye on the Olympic Games. Promising young athletes often have opportunities to progress rapidly in rigorous training programs. In some cases athletes are encouraged not only to perform well and perhaps set a new world’s record, but to win at any cost. A modern problem in the sports world is athletes’ use of drugs to enhance their performance and increase their chances of winning. Laws have been created to ban the use of such drugs.

3. For some people, having enough money to invest seems to be only a dream, since it is all they can just to make ends meet. A family may have enough to live on, enough to meet their needs, but they might have to struggle to pay off their debts. In this kind of situation it is necessary to plan expenditures carefully and keep to a budget, always looking for ways to cut corners. If a person earns a decent living and is able to follow a budget, even an average paycheck can go a long way.

II. Comprehension questions:

1. What problems can be encountered by a manager who can’t delegate responsibility?
2. What are some of the concerns people have about professional sports?
3. What are some of the steps in achieving financial security?

The pre-test / post-test for the flipped instruction

I. Paragraph reading:

1. Most time management plans include an orderly desk and desktop for doing all paperwork and information processing, with a well-organized filing system for hard copies of letters, bills, and so on, and orderly file folders on the computer for electronic copies of the same. The desk drawers are kept tidy so that nothing gets misplaced, and the top and desktop are kept clear in order to eliminate distractions. Ideally, this orderliness is reflected in the rest of the house, where there is “a place for everything, and everything in its place.”

2. Some people remember childhood as being a rather lonely time, with nobody to play with; they may have invented imaginary playmates to take the place of real ones. Most children, though, remember having “best friends” with whom they always played, as well as a group of neighborhood kids which functioned like a small community: some kids were looked up to, some get along with everyone, and a few never seemed to fit in. childhood experiences leave permanent impressions on people, and more than one unhappy adult remembers being made fun of by other children, teased with a nickname like “Fatty” or “Four eyes,” or always getting left out of games.

3. There is a downside to all this e-commerce, however. Pop-up ads, banners, and junk mail (also called spam) are the Internet’s version of advertising, and just as commercials interrupt television shows, so do advertisements interrupt the surfer’s leisure. For some people, closing the pop-up windows and deleting all the spams is an annoyance that takes up much of the time they should be enjoying. Firewalls and spam
filters are programs that help surfers to maintain security, which is a constant worry because of viruses, hackers, spyware, and identity theft.

II. Comprehension questions:

1. According to the time management experts, what are some time-saving techniques?
2. What are some games in which children use their imaginations?
3. What are some of the problems caused by the Internet?

Appendix B

Community of Inquiry Survey

This survey is to understand what your overall learning experiences were in the current instruction you have taken. There is no right or wrong answer. Please circle the answer which best reflects your overall thoughts about each statement. Your answers are ANONYMOUS and CONFIDENTIAL. Thank you in advance for your time.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

1. In the current instruction, the instructor clearly communicated important course topics. 1 2 3 4 5
2. In the current instruction, the instructor clearly communicated important course goals. 1 2 3 4 5
3. In the current instruction, the instructor provided clear instructions on how to participate in course learning activities. 1 2 3 4 5
4. In the current instruction, the instructor clearly communicated important due dates/time frames for learning activities. 1 2 3 4 5
5. In the current instruction, the instructor was helpful in identifying areas of agreement and disagreement on course topics that helped me to learn. 1 2 3 4 5
6. In the current instruction, the instructor was helpful in guiding the class towards understanding course topics in a way that helped me clarify my thinking. 1 2 3 4 5
7. In the current instruction, the instructor helped to keep course participants engaged and participating in productive dialogue. 1 2 3 4 5
8. In the current instruction, the instructor helped keep the course participants on task in a way that helped me to learn. 1 2 3 4 5
9. In the current instruction, the instructor encouraged course participants to explore new concepts in this course. 1 2 3 4 5
10. In the current instruction, instructor actions reinforced the development of a sense of community among course participants. 1 2 3 4 5
11. In the current instruction, the instructor helped to focus discussion on relevant issues in a way that helped me to learn. 1 2 3 4 5
12. In the current instruction, the instructor provided feedback that helped me understand my strengths and weaknesses relative to the course’s goals and objectives. 1 2 3 4 5
13. In the current instruction, the instructor provided feedback in a timely fashion. 1 2 3 4 5
14. Getting to know other course participants gave me a sense of belonging in the current instruction. 1 2 3 4 5
15. In the current instruction, I was able to form distinct impressions of some course participants. 1 2 3 4 5
16. Classroom communication is an excellent medium for social interaction. 1 2 3 4 5
17. I felt comfortable conversing through the class conversations. 1 2 3 4 5
18. I felt comfortable participating in the course discussions. 1 2 3 4 5
19. I felt comfortable interacting with other course participants. 1 2 3 4 5
20. I felt comfortable disagreeing with other course participants while still maintaining a sense of trust. 1 2 3 4 5
21. I felt that my point of view was acknowledged by other course participants. 1 2 3 4 5
22. Class discussions help me to develop a sense of collaboration. 1 2 3 4 5
23. Problems posed increased my interest in course issues. 1 2 3 4 5
24. In the current instruction, course activities piqued my curiosity. 1 2 3 4 5
25. In the current instruction, I felt motivated to explore content related questions. 1 2 3 4 5
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>I utilized a variety of information sources to explore problems posed in the current instruction.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>27</td>
<td>In the current class, brainstorming and finding relevant information helped me resolve content related questions.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>28</td>
<td>Class discussions were valuable in helping me appreciate different perspectives.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>29</td>
<td>In the current instruction, combining new information helped me answer questions raised in course activities.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>30</td>
<td>In the current instruction, learning activities helped me construct explanations/solutions.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>31</td>
<td>Reflection on course content and discussions helped me understand fundamental concepts in the current instruction.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>32</td>
<td>I can describe ways to test and apply the knowledge created in the current instruction.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>33</td>
<td>In the current instruction, I have developed solutions to course problems that can be applied in practice.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>34</td>
<td>I can apply the knowledge created in the current instruction to my work or other non-class related activities.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Guest Editorial: Learning Analytics in Technology Enhanced Language Learning

Yu-Ju Lan¹, Nian-Shing Chen² and Yao-Ting Sung³

¹Department of Applied Chinese Language and Culture, National Taiwan Normal University, Taiwan // ²Department of Information Management, National Sun Yat-sen University, Taiwan // ³Department of Educational Psychology and Counseling, National Taiwan Normal University, Taiwan // yujulan@gmail.com // nschen@mis.nsusu.edu.tw // sungtc@ntnu.edu.tw

Introduction

In the new era, language learning has become a lifelong issue. From primary school to college and workplace settings, a good command of a foreign language (FL) is highly valued as a requisite for participating competitively in the international community. To echo the trends featured in global FL learning, multiple modalities of technology enhanced language learning (TELL) are offered to satisfy diverse needs of FL learners. The various modalities can be briefly classified based on different perspectives as shown below:

- Formality: formal, informal
- Timeliness: synchronous, asynchronous
- Reality: real world, virtual reality, augmented reality, mixed reality
- Mobility: mobile, ubiquitous, specific location
- Openness: regular class schedule, MOOCs, open courseware

Unquestionably, no matter which perspectives are adopted, competency acquisition of the target language should be the main focus in FL learning. Furthermore, three components in successful FL learning are essential and cannot be omitted: learners, goals, and contexts. Obviously, the information that can assist the FL learners, educators, and researchers in meaningfully taking the above-mentioned components into account simultaneously would be of critical importance to successful FL learning.

As is known, specific characteristics of each individual, such as learners’ ages, learning styles, nationalities, motivation, learning goals, the experience in foreign language learning, learning behaviors, learners’ metacognition (self-regulation, self-estimation, etc.), and available learning time, are what may influence how well a second language can be acquired. Therefore, language educators and researchers definitely need the information to help them take as many learner variables as possible into account for providing FL learners with successful learning opportunities in all possible TELL modalities (Lan, 2009).

In terms of the goal component, learners and educators/researchers should always be discussed at the same time since they are the two sides of FL education. For FL educators/researchers, the information for assisting them to decide whether the goals of pedagogical syllabus are reached is as important as that for the FL learners to perceptual about their success/failure in FL learning. The techniques of assessment interpretation for supporting decision making by educators/researchers and learners, in the meanwhile, are also important and should definitely be valued whichever the TELL modalities are adopted by the educators/researchers or learners.

The last component, contexts, can be viewed as all the perceived phenomena including the physical surroundings in which a language is used. It can be the learning platforms/systems or environments (real or virtual) in which FL learners receive input or produce output of the target language. Additionally, language input from the environment, including contextual and non-linguistic cues, is easy to be comprehended by an L2 learner because he or she is in a low stress situation (Ray, 2012). Consequently, context-based FL learning is valued and emphasized in the issue of FL education and research in recent years (Lan, Chen, Li, & Grant, 2015; Lan, Kan, Sung, & Chang, 2016; Lan & Lin, 2016). However, the increasingly authentic learning environment makes it more challenging for FL educators/researchers to collect and evaluate students’ learning process, achievement, or behavior. Consequently, it has become a must to develop a technique to provide both educators/researchers and learners with timely information and clear explanation about FL learners’ learning log for improving the outcome of FL teaching and learning.

As described at the beginning, there are various options of the modalities which fit the FL learners’ learning needs the best. At the same time, the complete learning process of the FL learners recorded in the digital era, whichever the modalities are adopted, mostly becomes the BIG DATA. Consequently, the huge volume of data produced by the learners are so BIG that they cannot be easily handled and interpreted via the traditional
approach to yield meaningful and critical information related to the above-mentioned three essential components in FL learning for learners, educators, and researchers. Unfortunately, an overwhelming amount of information does not satisfy the needs of FL learners, educators, and researchers. In a word, developing more advanced techniques to better address FL learning is an important and urgent research issue in FL learning and teaching. Using learning analytics is considered as a common way to deal with the situations mentioned above and that is why this journal calls for this special issue.

Learning analytics refers to the technique to analyze the existing, learner-produced data for assessing academic progress, predicting future performance, giving suggestions, and spotting potential issues (Ali, Hatala, Gašević, & Jovanović, 2012; Duval, 2011; Johnson, Adams, & Cummins, 2012; Xing, Guo, Petakovic, & Goggins, 2015). With learning analytics techniques, educators are able to better satisfy L2 learners’ needs, predict L2 learning behaviors and outcomes, and provide L2 learners with personalized and adaptive learning (Godwin-Jones, 2014). Additionally, through data visualization, L2 learners, educators, and researchers can be better informed with timely decision-making information for improving their learning and teaching practices (Kickmeier-Rust, Bull, & Meissl-Egghart, 2014).

The aim of this special issue is to provide a platform for researchers to present their study efforts that may offer insights into the potential of using learning analytics to analyze language learning in different modalities and scenarios. These are open questions worth further exploration. Through the publication of this special issue, we can help develop a further understanding of the potential of learning analytics in TELL. After a rigorous review process, nine high-quality research papers have been accepted for publication in this special issue, and these papers clearly explain how learning analytics can be adopted in TELL to provide the learners, educators, and the researchers with insights into the real story of language learning from different perspectives. We hope that these studies will inspire future research in this direction.

In the first paper entitled “Visualization analytics for second language vocabulary learning in virtual worlds,” Hsiao, Lan, Kao, and Li developed a visualization analytic method based on the analysis on social network to examine the recorded learner paths within a virtual world during the learning process. They collected and analyzed the learning data of 14 college students from the virtual worlds while learning Mandarin Chinese vocabulary. Through the visualization analysis, the current study revealed a link among the learning paths, strategies and learners’ outcomes. The second paper with the title of “To activate English learning: Listen and speak in real life context with an AR featured u-learning system,” written by Ho, Hsieh, Sun, and Chen, presented a ubiquitous learning instruction system with augment reality features (UL-IAR). The aim was to improve the performance of EFL learning with authentic situations. Through a field experiment, the authors found that the learning outcomes were affected by students’ learning strategies and cognitive styles while students were using UL-IAR. That is, the enforcing learning strategy is more suitable for individuals with a field-dependent cognitive style than for other users who are field-independent and with mixed-field cognitive styles. Further, Pan investigated whether the proposed Kinect motion-sensing interactive system (KMIS) enhanced students’ English vocabulary learning. The paper “The effects of using the Kinect motion-sensing interactive system to Enhance English learning for elementary students” confirmed the effects of interactive games with a questioning strategy on students’ long-term retention of English vocabulary.

Next, in the fourth paper, Berger, Crossley, and Kyle introduced a model of lexical proficiency based on novel computational indices related to word context. By analyzing the correlations between lexical proficiency scores received from trained human raters and contextual indices, four indices related to associative, lexical, and semantic operationalizations of word context were confirmed. They found that computational measures of word context can predict human ratings of lexical proficiency and suggested that lexical, semantic, and associative context each plays an important role in the development of lexical proficiency. The fifth paper by Mørch, Engeness, Cheng, Cheung, and Wong evaluated the effects of the feedback provided by a writing aid system, EssayCritic, on essays written by students of English as a foreign language. By comparing two feedback conditions, one given by EssayCritic, and the other by peers, the authors found that the students in the EssayCritic group put more emphasis on the organization of their ideas while those in the contrary group included more ideas (contents) in their essays. Next in the sixth paper, “Assessing the language of chat for teamwork dialogue,” Shibani, Koh, Lai, and Shim focused on analyzing teamwork dialogue from a dataset of online chat data. By evaluating pre-processing and classification methods, their study provides a direction for assessing the language of chats for teamwork dialogue and helps extend the work of technology enhanced language learning to academic competency as well as to communication aspect.

Dahlgberg, in the seventh paper entitled “A multivocal approach in the analysis of online dialogue in the language-focused classroom in higher education,” proposed and evaluated a multivocal approach to analyzing...
online dialogue. She used practical data of screen recording from a 40-hour online Italian course to investigate the importance of an environment to the organization of the interaction among individuals, specifically the participants’ focus during the encounters. In the next paper, Chen and Yeh investigated the effects of cognitive styles on the use of hints in the context of academic English. Two types of hints were compared in their study, direct and indirect hints. By collecting and analyzing both quantitative and qualitative data, they confirmed that cognitive styles have considerable influence on students’ learning patterns in the context of academic English. Last but not least, the ninth paper by Wu and Huang developed a mobile game-based English vocabulary practice system aiming at enhancing students’ learning of English vocabulary. The proposed system entails selecting words according to textbook passages, a difficulty ratio, and learning portfolios. The practical evidence approved the effects of the proposed system on students’ English learning by different constructs.

Various learning analytics approaches were adopted in the abovementioned papers to analyze the data obtained from different TELL applications and therefore these papers provide insights into research outcomes from multi-perspectives. Thus, the papers included in this special issue will likely provide readers with a deep and extensive understanding of the potential of using learning analytics in TELL research. More issues for future research are undoubtedly to be inspired by reading the articles of the special issue.

References


Visualization Analytics for Second Language Vocabulary Learning in Virtual Worlds

Indy Y. T. Hsiao¹, Yu-Ju Lan¹*, Chia-Ling Kao¹ and Ping Li²

¹Department of Applied Chinese Language and Culture, National Taiwan Normal University, Taiwan // 
²Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University, University Park, PA, USA // indymajere@gmail.com // yujulan@gmail.com // c.kao@oxon.org // pul8@psu.edu

*Corresponding author

ABSTRACT

Language learning occurring in authentic contexts has been shown to be more effective. Virtual worlds provide simulated contexts that have the necessary elements of authentic contexts for language learning, and as a result, many studies have adopted virtual worlds as a useful platform for language learning. However, few studies so far have examined the relationship between learning outcomes and learning paths and strategies inside a virtual world. This study was designed to fill this research gap. In order to understand the impact of different learning strategies on learning outcomes in a virtual world, a visualization analytic method was developed to examine the recorded learner paths within a virtual world while learning occurred. In particular, the visualization analytic method adopted in this study was based on social network analysis. This study included 14 participants who were learners of Mandarin Chinese as a foreign language. The learning outcomes were based on their test scores from 7 learning sessions and the post-test. Through the visualization analysis, the current study revealed a link between the learning paths and strategies and learners’ outcomes. The strategies include the “nearest strategy,” the “focus strategy,” and the “cluster strategy.” The findings show that high-achieving and low-achieving students tend to use different strategies in learning new words. The visualization analytics thus effectively displays the learning strategies of vocabulary acquisition. Our methods could be applied to other second language learning studies, and the results can also provide insights into the construction of future virtual worlds for learning second languages.

Keywords

Visualization analytics, Learning analytics, Learning strategies, Virtual worlds, Second language learning, Vocabulary acquisition

Introduction

Language skills play a critical role in our globalized world for a person’s career development. A foundation of language learning is the acquisition of vocabulary (Barcroft, 2004; Lan, 2013; Llach & Gómez, 2007). Learning success in the acquisition of vocabulary will affect a student’s future skills in language learning in general, including, for example, reading skills (Cobb, 2007; Kern, 1989) and listening skills (Kelly, 1991; Smidt & Hegelheimer, 2004).

Many researchers (Krashen, 1985; Lan, Kan, Hsiao, Yang, & Chang, 2013) have pointed out the importance of learning a language in authentic contexts for faster and more effective acquisition. Several studies also stated learning in contexts facilitates learners’ acquisition of vocabulary (e.g., Snow, 2005). However, authentic contexts in the real world often involve travel and other time-consuming efforts. Virtual worlds provide an alternative to real world environments. Complex and perceptually rich environments in the virtual worlds can increase the student’s sense of immersion just as real environments do (Warburton, 2009). Lan (2015) discussed that in order to build an authentic and immersive learning environment, the government in Taipei funded 12 elementary schools to construct the English Village for language learning. However, such a program is very costly and still fails to meet all the needs of all the elementary school students. If Second Life is adopted as an alternative to build virtual environments for learning, positive results can be obtained to benefit all students (see Lan, 2015). Dalgarno and Lee (2010) also suggested that virtual worlds provide the third dimension via 3D virtual reality, and enhance the visual authenticity. Among the 3D virtual worlds available, Second Life (henceforth SL; Linden Research Inc.) has high flexibility and expandability (Kaplan & Haenlein, 2009), and has been very commonly used for teaching and learning. As an educator or a teacher, one can use SL to design the virtual environments in a manner required by the needs of a class.
Vocabulary acquisition in virtual worlds

In the context of second language learning, the acquisition of vocabulary is considered the most fundamental (Barcroft, 2004; Lan, 2013; Ilach & Gómez, 2007) because it affects how well a learner could perform in listening, speaking, reading, and writing. In order to carry out a conversation, a learner must acquire a certain amount of vocabulary and have an awareness of how each word could be appropriately used. However, the acquisition of vocabulary can be challenging (Lan, 2013; Li, 2015; Meara, 1980), given the large amount of words in a foreign language and the complex meaning-to-form mappings (Malt, Li, Pavlenko, Zhu, & Ameel, 2015; Zinszer, Malt, Ameel, & Li, 2014). To date, a large number of studies have examined how to best enhance the learning outcomes of vocabulary acquisition.

With the increasing use of virtual worlds, more researchers have been studying vocabulary learning in 3D virtual worlds. One reason could be that context-based learning allows L2 learners to build a direct link between vocabulary and its concepts, which facilitate L2 vocabulary learning (Cohen & Aphek, 1980; Kroll & Curley, 1988; Snow, 2005). Thus, many researchers have found that students who learn in virtual worlds receive positive outcome in vocabulary acquisition (Lan, 2015; Lan, Fang, Legault, & Li, 2015; Rankin, Gold, & Gooch, 2006).

For example, Lan et al. (2015) divided students into two groups, one learning vocabulary words through structured programs on computers, while the other through virtual worlds. Their findings showed that although participants were exposed to the same words for fewer times in the 3D virtual worlds than in computer-assisted learning, the two groups performed equally well at the end of the training. However, the researchers did not use additional data analytics for understanding participants’ language learning strategies (LLSs). As Lan (2013) pointed, data analytics on LLSs could help reveal additional insights into language learning. The current study is designed to examine the LLSs used for vocabulary learning by the same participants reported by Lan et al. (2015).

Analysis of language learning strategies (LLSs) in virtual worlds

Language Learning Strategies or LLSs are defined as “specifications, behaviors, steps, or techniques -- such as seeking out conversation partners, or giving oneself encouragement to tackle a difficult language task -- used by students to enhance their own learning” (Scarcella & Oxford, 1992, p. 63). LLSs can be divided into six types: cognitive, metacognitive, memory-related, compensatory, affective, and social, according to Oxford (2003). On the basis of Oxford’s (1990) research on LLSs, Lan (2013) further outlined 12 vocabulary learning strategies: practicing, note-taking, key word, contextualization, grouping, imagery, recombination, deduction, analysis, physical response, translation, and transfer. In this study, an LLSs questionnaire based on Lan’s (2013) 12 learning strategies was used to analyze the participants’ learning data and to identify the LLSs used by the participants.

As Saggarra and Alba (2006) stated, how students learn vocabulary words influences how well they understand and use the words. Although previous studies indicated that learning outcomes are linked to students’ learning logs, which cover all of the students’ learning activities (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Lan et al., 2013; Rankin et al., 2006), they focused on the interactions among different users in a digital learning environment. The current research, however, highlights the interactions between the student learners and the target vocabulary words.

Various methods can be used to analyze videotapes, questionnaires, interviews, and the database of learning logs displayed in pie charts, line graphs, etc. (Heer, Bostock, & Ogievetsky, 2010). In past research, questionnaire has been the most commonly used form for evaluating performance of vocabulary learning (Gu & Johnson, 1996; Kojic-Sabo & Lightbown, 1999). However, few studies have recorded detailed learning logs. The limitation is that learners themselves could not report everything in details during the learning process, and therefore, these methods cannot provide an in-depth understanding of how learners actually learn in virtual worlds. At the same time, simply relying on databases based on virtual words is not the best choice because 3D virtual worlds are much more complex than 2D environments. For example, Cruz-Benito et al. (2014) provided the teachers and students with learning analytics displayed in texts and statistics; Camilleri, de Freitas, Montebello and McDonagh-Smith (2013) visualized the interactions in virtual worlds through statistical graphs; Tashiro, Hung and Martin (2010) visualized the learning paths of individual learners. These studies are useful, but none of them captured the huge amount of data from students’ learning behaviors. Although it is possible to manually analyze learning logs, e.g., how students interact with one another and how students interact with virtual learning objects, such manual methods are only suitable for handling a small portion of the big data. Furthermore, it can be time-consuming and resource-demanding. To solve this problem, big data analytic tools are required.
With the rapid development of big data analytics, tools and methods have emerged to examine the learning data from virtual worlds. However, these analyses so far have been mostly developed for analyzing 2D websites. For example, Minović and Milovanović (2013) produced a platform which provides educators with the ability to define a 2D adventure game, and with graphical representation of what students know. This technique is also used to assist the management of virtual worlds (Perera, Allison, Ajinomoh, & Miller, 2012), but it is rarely used to analyze the users’ strategies in virtual worlds. In the study we follow the approach used in several recent studies to deal with big data in e-learning. For example, Bargel, Schröck, Szentes, and Roller (2012) used visualization technique to study the learning paths of students taking online courses; Lin, Yeh, Hung, and Chang (2013), for optimizing performance of creativity, used data-mining techniques of decision trees to provide personalized learning paths in a web-based learning environment. These are encouraging studies, but they are not methods that can be directly used to examine learning paths that occur in 3D virtual worlds.

In the current study, a learning path is defined as the chosen route of navigation in a virtual learning environment which enables the learner to build the target knowledge progressively. As Walkowiak, Foulsham and Eardley (2015) showed, learners’ navigational strategies can affect which learning paths they take. The strategy for deciding learning paths is called the navigation strategy, which may influence performance on independent navigation tasks (Rodgers, Sindone III, & Moffat, 2011). Navigational strategy involves how users want to shift between locations and how they want to operate a virtual vehicle (Martens & Antonenko, 2012). More specifically, the use of effective navigation can affect the learners’ degree of satisfaction as well as the learning outcomes (Martens & Antonenko, 2012). In virtual worlds, the acquisition of vocabulary can be viewed as a navigation learning task; thus, navigation strategy is considered as a type of LLSs and it tends to influence learning outcomes (Gu & Johnson, 1996).

In sum, the current study aims at examining the LLSs in an attempt to understand the role of individual differences in second language vocabulary acquisition. Specifically, using visualization analytic tools, the authors would identify student learning paths to better understand the LLSs that may be related to learning outcomes. The key research questions are: (1) What LLSs do learners adopt in the virtual worlds while learning a new vocabulary word? (2) How do the LLSs influence learners’ performance in the acquisition of vocabulary in virtual worlds?

**Method**

**Participants**

Fourteen students aged between 19 and 28 (mean age = 22; 8 females) from the Pennsylvania State University participated in the study. They completed the Language History Questionnaire (Li, Zhang, Tsai, & Puls, 2014), used to understand their language learning background and language proficiency level before the study, and the results showed that all the learners were native speakers of English and had no prior experience with Mandarin Chinese as a foreign language. See Lan et al. (2015) for details. The study of Lan et al. (2015) also contained another group of 17 students who learned Mandarin Chinese in a computer-assisted picture-word association paradigm, and data from those participants are not included in the current analyses, as those data are not subject to the visualization analytics presented here.

**Instruments**

*Learning materials*

Three virtual world contexts were used in this study: a supermarket, a kitchen, and a zoo. Figure 1 presents an example of the virtual zoo. Learners were required to visit all three contexts to learn 90 new Chinese words, 30 words in each context. In the virtual world they were free to take any learning paths within the time limit (see below), and can click on any virtual object to learn its Chinese name. Once the participants clicked the objects, they would hear the pronunciations of the corresponding words. All words to be learned were two-syllable Chinese words. The participants were instructed that the objects labelled with “TELL” in blue (above the objects) were the target words to be learned, because there were other objects in the virtual words that were not the learning targets. Once an object was clicked, its label would turn red, indicating that it had been learned at least once. The learner could click on the same object for as many times as needed, within the 5-min time window. See Lan et al. (2015, pp. 675-676) for further details.
Figure 1. The virtual zoo for L2 word learning; the learner avatar can move around in the zoo and tap any of the 30 animals to listen to the corresponding L2 term/word referring to the animal

Learning logs

For recording and analyzing the digital learning data, the chosen tools were LSL (Linden Script Language), PHP, and MySQL to develop the Second Life Learning Database (SLLDB), which recorded all learners’ event logs, including learner information, words learned, learning dates, time, and the 3D locations of the objects/words learned (The SLLDB is independent of the Second Life platform and was developed by the authors before the participants took part in the study). In addition, the authors transformed the learning logs into data that could be read by the visualization tools used in this study.

Interview

After the completions of all the learning and testing sessions, the participants were required to participate in an interview so that the authors could explore the participants’ learning strategies. The interview contained 12 questions, and was compiled based on the 12 vocabulary learning strategies proposed by Lan (2013). The complete Word Learning Strategy Questionnaire can be found here (https://goo.gl/05Wy7v).

Procedure

The learning and testing sessions lasted 18 days, during which each participant attended 7 learning sessions, each consisting 3 learning contexts. Before the experiment, all participants completed the informed consent approved by the Institutional Review Board of the Pennsylvania State University. They then did a pre-test with practice trials. They made their own avatars walk around the virtual environments and clicked on different objects (which were similar but not identical to the objects used in the experimental sessions). The practice session included fewer words and a shorter duration (3 minutes). During the experimental sessions, each participant was given 5 minutes to learn 30 vocabulary words in each of the three learning contexts (see Learning materials), and there was a 3-min break between two learning contexts. The total duration of a session was 21 minutes. Each learning session was followed by a test to evaluate student’s learning outcome. During the test, participants were required to listen to audio files and to select the correct images corresponding to the audio files. All the images were screenshots of the objects in the three SL virtual learning contexts corresponding to the 90 target words required for learning. Information on how they answered each question, including the length of time taken to answer a question and their accuracy, was recorded by the E-Prime software (Schneider, Eschman, & Zuccolotto, 2012). See Lan et al. (2015; pp. 676-677) for further details. The complete procedure of each learning and testing sessions is displayed in Figure 2.

The steps from “Supermarket” to “Test” were repeated for 7 times in 3 weeks. Three weeks after the end of the 7 learning sessions, the participants were asked to do a post-test and an interview. Based on how the participants performed in the post-test, they were divided into the high-achieving group and the low-achieving group. The high-achieving group included 8 students scoring over 85% in accuracy in the post-test (mean: 93.6%), and among them, 7 of them received at least 80% in the 3rd test and 1 student scored 78% in the 3rd test and 94.44% in the 4th test. The low-achieving group scored less than 85% in the post-test (mean: 67%), and only one student reached over 80% in the 4th and the 5th tests.
**Data analysis**

All participants’ learning logs were saved in the SLLDB. Adjacency matrices and clusters produced by the learning paths were used to identify the commonly used learning strategies, the differences between the strategies used by the high-achieving and low-achieving students, and the possible explanation for such differences. The complete and detailed method used to compute the adjacency matrices and the clusters for each participant can be found here (https://goo.gl/R9l2QK). See Figure 3(a-b) for an example of adjacency matrix.

Due to the large amount of learning logs in virtual worlds, the R statistical software was adopted (R version 3.2.4; Development Core Team, 2004) and several modifications were made based on Lincoln’s (2014) research. With the data format employed in Lincoln’s research, the researchers compiled the nodes and edges. Three more algorithms were used: label propagation community, leading eigenvector community, and infomap community to analyze the data collected. The method of label propagation is a method based on Social Network Analysis (SNA; Otte & Rousseau, 2002). Although powerful, this method, however, cannot provide more details through adjacency matrices about the learning paths among nodes, such as the frequency and the directions of learning for specific words (Wong, Mackey, Foote, & May, 2013). Thus, to understand how students tended to select certain words for learning, Gephi (Bastian, Heymann, & Jacomy, 2009) was used to further observe the participants’ learning paths for the acquisition of vocabulary. Gephi is an open graph visualization platform (see https://gephi.org/). It shows the clusters in a network and the directional patterns and strength of any interrelationships. Gephi 0.91 (Bastian et al., 2009) was adopted and each learning object was treated as an individual in the network and the learning paths as the interrelationships. Compared to other layout algorithm, it produces high-quality results (Khokhar, 2015). It aims at exploring real-world data with high-quality visualization graphs in the output (see https://gephi.org/ for further details). More details about label propagation and Gephi can be found here (https://goo.gl/P3ol6D)

**Results**

**Learning strategies in the virtual world**

Figure 3(a-b) shows the adjacency matrices of objects in the virtual supermarket, the order of vocabulary words arranged according to where they were located in the virtual world, for low-achieving vs. high-achieving learners, respectively.
Figure 3(a). The adjacency matrices of objects in the virtual supermarket from low-achieving learners

Figure 3(b). The adjacency matrices of objects in the virtual supermarket from high-achieving learners

The different colors represent different clusters formed by the participants’ learning paths. For example, the green block at the bottom left in Figure 3(a) indicates a cluster with words 白菜 ("cabbage"), 玉米 ("corn"), and 芹菜 ("celery"). Since the order of the words was arranged according to how they were located in the virtual world, each colored block in Figure 3(a-b) indicates the words that the learners grouped together as being close. The size of the block indicates the number of words categorized into a single group, and the color of the block indicates a single group of words.
This result shows that the participants tended to select the words which were the nearest to the previous word learned. For the naming of the labels, the distance-based method is a common one used by many researchers. We employed the data analysis method used by Kirsten, Wrobel, and Horváth (2001), Kuncheva and Bezdek (1998), and Biswas and Robinson (2008) to analyze the data. The method “nearest neighbour” is considered instance-based learning or “lazy learning” because by using the nearest neighbors the students do not have to make active choices but to wait for the nearest object (Kirsten et al., 2001). In the current data, it is found that this “nearest strategy” was commonly used by many students.

Figure 3(a) shows that luabo (“carrot”) itself formed a group. There are two conditions under which a single word can form a group: (a) there is no any specific route connected between the target word and other words, and (b) the target word has been frequently clicked on for several times. Given these two conditions, a single-word group indicates a “focus strategy,” in contrast to the “nearest strategy.”

A third strategy identified is the “cluster strategy,” which means that the students selected the words that belong to the same group but are not nearby. For example, as shown in Figure 3(b), the word naiyou (“butter”) and mianbao (“bread”) belonged to the same group not because they were located nearby but because the student linked butter with bread functionally (i.e., people commonly put butter on bread). Further, the ‘cluster strategy’ was divided into categories based on the objects’ functions, appearances, pronunciations, and the combination of function and pronunciations of the virtual objects.

The remaining grey blocks and white areas shown in Figure 3(a-b) indicate no specific strategies among the learning paths for the words. That is, those were the words randomly picked by the students for learning, which is not considered as a strategy in this study.

**Different strategies used by high- versus low-achieving learners**

Based on the three strategies described above and the results shown in Figure 3(a-b), Table 1 summarizes the distribution of type of strategies adopted by the high- and the low-achieving learners.

<table>
<thead>
<tr>
<th>Type of strategy</th>
<th>Frequency</th>
<th>High-achieving learners</th>
<th>Low-achieving learners</th>
<th>Number of the same clusters by high- and low-achieving learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest strategy</td>
<td>47 (54.65%)</td>
<td>21 (24.42%)</td>
<td>26 (30.23%)</td>
<td>8 (18.60%)</td>
</tr>
<tr>
<td>Focus strategy</td>
<td>30 (34.88%)</td>
<td>20 (23.26%)</td>
<td>10 (11.63%)</td>
<td>7 (16.28%)</td>
</tr>
<tr>
<td>Cluster (Appearance)</td>
<td>4 (4.65%)</td>
<td>4 (4.65%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>Cluster (Function)</td>
<td>1 (1.16%)</td>
<td>1 (1.16%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>Cluster (Pronunciation)</td>
<td>2 (2.33%)</td>
<td>1 (1.16%)</td>
<td>1 (1.16%)</td>
<td>1 (2.33%)</td>
</tr>
<tr>
<td>Cluster (Function and Pronunciation)</td>
<td>2 (2.33%)</td>
<td>2 (2.33%)</td>
<td>0 (0.00%)</td>
<td>0 (0.00%)</td>
</tr>
<tr>
<td>Total</td>
<td>86 (100%)</td>
<td>49 (56.98%)</td>
<td>37 (43.02%)</td>
<td>16 (37.21%)</td>
</tr>
</tbody>
</table>

Table 1 shows that the most frequently used strategy is the “nearest strategy,” occupying 54.65% of all the strategies used. The percentage of the “focus strategy” is 34.88%. The frequency of the use of this strategy is twice as high in the high-achieving learners (23.26%) as in the low-achieving learners (11.63%). A chi-square test revealed a significant difference in the type of strategy adopted by the two groups of students ($p < .05$; $df = 2$).

Additionally, the two groups also differed in the adoption of the “cluster strategy”: high-achieving learners used 9.30% and low-achieving learners used 1.16% of this strategy. The detailed results, which can be found here (https://goo.gl/wVniXx), show that groupings created by the high-achieving students were more meaningful. Kojic-Sabo and Lightbown (1999) also adopted the cluster analysis method, but their data were based on the results of questionnaires and test scores. However, the analyses based on the visualization analytic methods are consistent with the basic patterns of Kojic-Sabo et al. (1999), showing that learning strategies are clearly connected with learning outcomes.

With a further analysis through set color by edge weight, the results can be seen in Figure 4(a-b). The horizontal x-axis shows the starting point, while the vertical y-axis shows the destination point. The colored blocks indicate where the starting and the ending points meet, showing the learning paths between different words. Warm colors
are red and yellow, and cool colors are blue and green. The warmer the colors of the blocks, the more learning paths there are between the two words. As displayed in Figure 4(a-b), the high-achieving learners tended to learn vocabulary words through repetitive clicks on the same objects. For example, as Figure 4(b) shows, the color of the block for 茄子 qiezi (“eggplant”) to 黄瓜 huanggua (“cucumber”) is as warm as the color of the block for 茄子 qiezi (“eggplant”) to 茄子 qiezi (“eggplant”). By contrast, in Figure 4(a), for the low-achieving students, the color of the block for 茄子 qiezi (“eggplant”) to 黄瓜 huanggua (“cucumber”) is warmer than the color of the block for 茄子 qiezi (“eggplant”) to 茄子 qiezi (“eggplant”).

**Figure 4(a).** Set color of objects in the virtual supermarket used by low-achieving learners

**Figure 4(b).** Set color of objects in the virtual supermarket used by high-achieving learners
Visualization analyses of high- versus low-achieving learners

Although the user strategies could be identified through adjacency matrices as discussed above, the complete learning logs could be used to provide further detailed analyses of individual differences between groups. An SQL script was created to link the MySQL database of SLLDB with Gephi to visualize the data. Take for example the results of the visualization graphs produced by Gephi through the force atlas layout, as shown in Figure 5(a-b), they are the learning logs of the high- versus low-achieving learners in the virtual kitchen. Here the researchers illustrate the analysis with the data from the virtual kitchen. Although the results from other virtual contexts are not displayed here, they are consistent with the following analysis based on the data from the virtual kitchen.

Figure 5(a). Gephi displays of objects in the virtual kitchen used by low-achieving learners

Figure 5(b). Gephi displays of objects in the virtual kitchen used by high-achieving learners

Figure 5(a-b) shows the visualization graph produced by Gephi through the force atlas layout. All of the learning paths and their directions are shown in lines with arrows. The thickness of lines indicates the frequency of such learning paths. The force atlas could set the repulsion strength, attraction strength, and gravity according to the size of words and thickness of lines, and the position of words and the distance between words are then adjusted. It is important to note here that the thickness does not represent an absolute value. Rather, the thickness was
decided upon the frequency of certain learning paths within the group. That is, the different levels of thickness presented in Figure 5(a) and (b) should not be directly compared.

The force atlas could also distribute the attractive force along outbound links. Thus, words would be pushed away if their attraction strength is lower than that of their neighbors. In this way, the clusters of words would be clearer. As can be seen, the number of clusters created by the high-achieving group was greater than the number of clusters created by the low-achieving students, and the clusters are also clearer in the former than they are in the latter: the average clustering coefficient of high-achieving students is 0.759, while that of the low-achieving students is 0.718. In other words, the high-achieving students were more able to tell the difference between different words and create more meaningful links, while the learning paths created by the low-achieving students were quite evenly distributed, indicating that for them, most words were similar in features. Figure 5(b) also shows that high-achieving students created much thicker lines with two-way arrows, and more meaningful clusters of words were formed since only those words with a greater number of interactions in between could form clusters.

Further, the colored lines indicate the learning paths of the back-and-forth paths between two words. For example, the green lines shown in Figure 5(a-b) indicate that learners’ repeated learning of those two words twice, while the blue lines indicate learning three times. The fact that more blue and green lines can be found in Figure 5(b) means the high-achieving learners tended to click between words more often.

Finally, the size of the words reflected the number of times the word was clicked on, with bigger fonts indicating more frequent clicks/learning by the learner. Similar to the description of thickness, the size of the words does not represent an absolute value and is based on the results within a single group; thus, the sizes cannot be directly compared between Figure 5(a) and (b). The results show that the low-achieving group tended to click on objects which were more easily seen. For example, the sizes of 蒸锅 zhengguo (“steamer”), 蒸锅 zhengguo (“steamer”), and 烤箱 kaoxiang (“oven”) in Figure 5(a) are bigger. This could be because these virtual objects were relatively larger than all the other objects and could be more easily seen and clicked on. In addition, the words next to 餐桌 canzhuo (“table”) also received more frequent clicks. On the other hand, the high-achieving group clicked on virtual objects more selectively than did the low-achieving group. Specifically, the high-achieving learners did not simply click frequently on a group of words that are close in physical distance in the virtual environment.

The current study, through the visualization of Gephi, not only imported and analyzed the information on the source, target, and edge, but also the information on the number of repeated learning paths and the person who created such paths. Thus, by using the filter function of Gephi (https://gephi.org/tutorials/gephi-tutorial-quick_start.pdf), this study has obtained specific results according to specific conditions in this regard. Figure 6(a-b) shows the results of the learning paths that had more than 3 times of back and forth on the same pairs of words.

As can be seen in Figure 6(a-b), the high-achieving students created much more meaningful clusters. For example, as Figure 6(b) shows, Student #3032 (All student learners were numbered by random ID numbers) from the high-achieving group repeated the clicks on the pair of 椅子 yizi (“chair”) and 凳子 dengzi (“stool”) and the pair of 茶壶 chahu (“teapot”) and 茶杯 chabei (“teacup”), indicating that this student applied the “cluster strategy” of pronunciation over 3 times (perhaps because of the similar syllables involved in the two pairs of word). Similarly, Student #3039 also used the “cluster strategy” of pronunciation by repeatedly clicking on 蒸锅 zhengguo (“steamer”) and 烤箱 kaoxiang (“oven”) over 3 times. In contrast, Figure 6(a) shows that while Student #3036 showed similar patterns as students from the high-achieving group for 蒸锅 zhengguo (“steamer”) and 烤箱 kaoxiang (“oven”), Student #3045 repeatedly clicked on the pair of 手套 shoutao (“mitten”) and 烤箱 kaoxiang (“oven”) and the pair of 水槽 shuicao (“sink”) and 烤箱 kaoxiang (“oven”). Since the latter pair of words were located next to each other, it implies that Student #3045 was able to link two words together simply based on the position of objects. A more in-depth analysis of the interview also revealed similar results. The detailed and further explanation of the results can be found here (https://goo.gl/GJHKg6).

Although some findings are not significant enough due to the small sample size, these visualization analyses provide informative data about students’ learning behaviors, and they could also serve as useful reference for instruction, for example, to the teachers who want to know what words the students are focusing on in learning. Meaningful repetition of words is an important indicator of whether the learner uses strategic learning. The results show that 5.14% of the learning paths created by the high-achieving students can be regarded as “repetition,” while the percentage of “repetition” created by the low-achieving students is 3.72%. Out of 36,734
repetitions, 99% occurred below 3 times, while 27 repetitions were found to occur over 3 times. Among the word pairs that had occurrences over 3 times, the repetitive clicks between 松鼠 songshu ("squirrel") and 兔子 tuzi ("rabbit") (cluster by appearance) reached as many as 6 times by the high-achieving students, and those between 大象 daxiang ("elephant") and 鴕鳥 tuoniao ("ostrich") as many as 11 times by the low-achieving student. The repetitions created by the high-achieving students are considered more meaningful because they are based on similarity of either the pronunciation or the appearance of the virtual objects.

Figure 6(a). Gephi displays of objects in the virtual kitchen repeatedly clicked for more than 3 times by low-achieving learners

Figure 6(b). Gephi displays of objects in the virtual kitchen repeatedly clicked for more than 3 times by high-achieving learners

When interviewed, the participants from different groups provided quite different answers. The main difference could be seen particularly in when asked whether they found any rules for Chinese words. All of the students in the high-achieving group answered “Yes,” while only half of the students in the low-achieving group answered “Yes.” This finding could indicate that the high-achieving students could identify either the link or the rules between the two words and that it may also explain why the students from the two groups adopted different learning strategies (cluster strategy by the high-achieving group vs. nearest strategy by the low-achieving group).
Discussion

The current study combines tools for data visualization and database from learning logs recorded from virtual worlds (SLLDB) to analyse learners’ behaviours while they learned vocabulary of Mandarin as a foreign language in a virtual world. It should be noted that the three contexts included in this research display different features in terms of the arrangement of the space. Thus, when presenting the results of the adjacency matrices, the authors especially selected the results obtained from the supermarket due to its “U”-shaped arrangement. Presumably, the learners would be likely to click on different words based on what’s close to the words previously learned, clearly one of the results obtained from the low-achieving students. However, the high-achieving students were able to conduct more strategic learning by creating different clusters not merely based on the physical positions of the objects in the virtual worlds. When presenting the results presented by Gephi, the authors chose the results received from the kitchen because students from both groups could easily create clusters, for the objects in the kitchen were arranged without a specific order, forcing the students to rely on specific LLSs to help decide their learning paths. The zoo was designed as a round area, in which students were more likely to walk around and along the ring. That’s why fewer clusters were formed because it would be inconvenient for students to click between objects back and forth. That being said, the results obtained from different contexts were consistent with each other regardless of which visualization analytics, R or Gephi, was used.

The visualization tools allowed us to effectively examine the links between students’ choice of learning strategies and their learning paths. The three main learning strategies are the “nearest strategy,” the “focus strategy,” and the “cluster strategy.” The “cluster strategy” is further divided into three types: by appearance, by pronunciation, and by function. In addition, the high-achieving learners tended to employ particular learning strategy to do vocabulary learning, while the low-achieving learners showed more tendency to learn words either by using the nearest neighbours strategy or randomly. The analyses indicated that there is a difference in the learning strategies adopted by the high- versus the low-achieving learners.

The findings also showed a few features from the students’ learning paths. Previous studies such as Camilleri et al. (2013) also provided visualization analysis results, which mostly included visit time and traffic. The work reported here demonstrates that the results based on social network analytic methods can provide further detailed information, including visualizing the links among different elements/words/objects. Other studies such as Tashiro et al. (2010) used PathFinder to visualize student learning paths, but their graphs show only the learning paths of individual students. The analyses can provide graphs of not only individual learners, but also different learning groups. With this type of graph analysis, researchers can effectively capture student learning in self-exploratory environments such as the SL virtual worlds, and further provide teachers with the whole picture of students’ behaviours as well as individual learning paths.

The visualization method used in this study provides useful information for designing learning environments that can more effectively guide their students in learning vocabulary or other aspects of languages. It is suggested that researchers design the environments based on the features of objects clustering, and make best use of the space and the arrangement of virtual objects in virtual worlds. For example, if the virtual environment can be arranged to place objects similar in appearances, function, or pronunciations in the same area, learners can then more easily navigate among the words that are similar and build learning paths that are shorter and more effective.

One limitation of the current exploratory work is the small number of participants and the short training/learning period for the results to be generalized to all students. Another limitation is that this research has focused on the learning of Mandarin, and whether similar patterns would be obtained for other languages learned in virtual worlds is yet unclear. Finally, it would also be interesting to find out whether the learner’s spatial ability would affect their performance in language learning in virtual worlds. Despite these limitations, it is hoped that this work will provide an initial effort to analyze language learning big data by employing the visualization analytics to identify learning paths, learning strategies, and their relations with learning outcomes.

Acknowledgments

We would like to thank two anonymous reviewers and editors for their valuable comments and suggestions for improving this article. We thank the Ministry of Science and Technology, Taiwan, R.O.C., under Grant Nos. MOST 103-2628-S-003 -002 -MY3, MOST 105-2511-S-003 -018 -MY3, MOST 105-2911-I-003-301, and the US National Science Foundation grants, BCS-1349110 and BCS-1338946, for financially supporting this
research. We are also grateful that this research was partially supported by the Aim for the Top University Project and Center of Learning Technology for Chinese of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C., and the International Research-Intensive Center of Excellence Program of NTNU.

References


To Activate English Learning: Listen and Speak in Real Life Context with an AR Featured U-Learning System

Shu-Chun Ho¹*, Sheng-Wen Hsieh², Pei-Chen Sun¹ and Cheng-Ming Chen¹

¹Department of Software Engineering and Management, National Kaohsiung Normal University, Taiwan // ²Department of Innovative Design and Entrepreneurship Management, Far East University, Taiwan // sch@nknu.edu.tw // onyxhsw@gmail.com // sun@nknu.edu.tw // tournesol231@gmail.com

*Corresponding author

ABSTRACT
The increasing advance of mobile devices and wireless technologies has generated great interest in ubiquitous learning (u-learning) among academia, practitioners, and policy makers. However, design elements that incorporate learning styles and learning strategies into u-learning system applications in English as a Foreign Language (EFL) education are still limited. There are two objectives in this research. First, we developed a Ubiquitous Learning Instruction System with Augment Reality features (UL-IAR) to improve the performance of EFL learning with authentic situations. Second, we examined whether different learning strategies and cognitive styles affect learning performance in using UL-IAR. We conducted field experiments to investigate the appropriation of learning strategies and cognitive styles in using UL-IAR to learn EFL. The results showed that learning strategies and users’ cognitive styles affect learning performance in using UL-IAR. Individuals with field dependent cognitive style fit enforcing learning strategy better than other users who are field independent and mix field cognitive styles. Our findings provide theoretical and practical insights for pedagogies that are suitable for u-learning. Our findings also contribute to the practice of AR and u-learning system development.

Keywords
Augmented Reality (AR), Cognitive Style (CS), Context-aware, Learning strategies, Ubiquitous learning

Introduction

Technological advancements in wireless applications and the popularity of mobile devices have created great advantages and opportunities for learning (Huang & Chiu, 2015; Hsieh et al., 2016). As a natural complement to ubiquitous computing, AR provides an intuitive and interactive interface to a three-dimensional information space embedded within physical reality. The application of Augmented Reality (AR) with context-aware ubiquitous learning (u-learning) systems offers considerable advantages for English as a foreign language (EFL) learning. EFL learners often confront difficulties and challenges when they need to speak English in real life settings. Contextual learning environments have been proven to improve performance in EFL learning (Sandberg et al., 2011). Context-aware ubiquitous learning (u-learning) refers to the computer systems that seamlessly integrate into our everyday life. It is a version of our future computing lifestyle in which providing services and information occurs anywhere and anytime. Researchers have found positive outcomes in the use of mobile devices for English language learners. Kolb (1984) emphasized learning is constructed and contextual which relies on an engagement with social interactions and experience from the real world (De Freitas & Neumann, 2009). Based on Kolb’s arguments, we intend to design and develop a U-Learning Instruction System with Augment Reality Features (UL-IAR) to help users to learn EFL in real life contexts. Our UL-IAR allows users to learn EFL in a constructed and contextual environment that involves social interactions and experience.

Despite prior studies have considered the technical affordances of u-learning, there is an ongoing need to examine pedagogies that are suitable for u-learning to inform teacher practice, policy makers, curriculum developers, and teacher education (Pegrum et al., 2013). To implement a context-aware u-learning system for EFL learners, we further design learning strategies for different individuals. We believe that individual difference also leads to different learning performance in using u-learning systems. The individual differentiation can be discerned through cognitive styles, learning styles, cognitive control, and intelligence (Jonassen & Grabowski, 1993). The design of learning systems must consider individuals’ cognitive styles which would increase the efficacy and satisfaction of the learning experience. However, the performance outcomes of distinct cognitive styles that applied to new learning technologies remain inconclusive (Dillon & Gabbard, 1998). This knowledge gap highlights the need to examine the difference between cognitive styles in u-learning systems with AR. Therefore, our research objective is to empirically examine the learning performance of u-learning system with different learning strategies and cognitive styles. Overall, this paper intends to answer the following research questions:

- What kind of learning strategy leads to better EFL learning performance when using UL-IAR?
- Do the users with different cognitive styles have different learning performance when using UL-IAR?
- Is there best alignment between learning strategies and cognitive styles when using UL-IAR?
This paper is organized as follows. We first review the definitions and studies of u-learning, augmented reality, and cognitive styles. We then show our research method which includes system design and development, experimental design and experiment participants. Last, we discuss our research results and findings and conclude theoretical and practical findings, limitations, and future research.

**Literature review**

This section reviews the definitions and studies of ubiquitous learning, the applications of augmented reality in education, and cognitive styles.

**Ubiquitous learning**

Ubiquitous learning (u-learning) refers to a ubiquitous learning environment that enables individuals to conduct learning activities in the right place and right time. U-learning environments create a context-aware, interoperable, pervasive, and interactive learning architecture that integrates, connects, and shares learning resources among the appropriate parties (Huang et al., 2011). Context-aware u-learning environments can better understand learners' activities and the timely geographical parameters in the real world such as the locations and behaviors of the learners. In other words, context-aware u-learning environments should support personalization and the situation of instructional activities (Chen & Li, 2010). It also supports collaborative and interactive learning (Peng et al., 2008). Moreover, it can achieve constructive learning and self-regulated learning through learning activities (Chen & Chung, 2008). Thus, u-learning is a learning paradigm that takes place in a ubiquitous computing environment that enables users to learn at the right place at the right time (Shih et al., 2012). In this study, we define the context-aware u-learning environment as an environment that users access through mobile devices, wireless communications, and sensor technologies when participating in learning activities.

Context-aware u-learning has been applied in different disciplines such as natural science (Hwang et al., 2010; Chu et al., 2010; Huang & Chiu, 2015; Shih et al., 2016), museum learning (Hall & Bannon, 2006), language learning (Lan & Lin, 2016), sharing of learning resources (Yu et al., 2015), and certification tutoring systems (Huang et al., 2016b). However, the inadequate application of novel technology may lead to meaningless learning of operational techniques, rather than practical, meaningful involvement in learning (Huang & Chiu, 2015). Despite the various studies that have been conducted to examine the technical affordances of u-learning, there is an ongoing need to examine pedagogies that are suitable for u-learning to inform teacher practice, policy makers, curriculum developers, and teacher education (Shih et al., 2012; Pegrum et al., 2013). It is necessary to evaluate whether a u-learning system is able to achieve the purpose of meaningful learning (Huang & Chiu, 2015). In addition, the predictions about transformational teaching practices seem promising but the widespread and effective application of u-learning has not been realized yet. Moreover, prior studies have suggested that evaluation is critical to improve the quality of technology-supported learning environments (Martínez-Torres et al., 2008; Huang & Chiu, 2015). We apply augmented reality (AR) in our context-aware u-learning system to assist English learning in authentic situations.

**Augmented reality**

Augmented reality (AR) refers to technologies that enhance the sense of reality enabling the concurrence of digital information and real environments (Azuma, 1997; Sayed et al., 2011; Chang et al., 2015b). An AR system is able to combine or supplement real world objects with virtual objects or superimposed information (Azuma, 1997; Bacca et al., 2014; Sayed et al., 2011; EDUCAUSE Learning Initiative, 2005; Chang et al., 2015b). Instead of replacing reality completely, AR intends to supplement reality (Azuma, 1997). AR is not only a virtual technique but also an imbedded design characteristic that helps users retrieve the proper information at the right time and right location (EDUCAUSE Learning Initiative, 2005). An AR system has three characteristics (Azuma, 1997). First, it combines real and virtual objects in the environment. Second, it allows real-time interaction. Third, it aligns real objects or places and digital information in 3D (Azuma et al., 2001). These characteristics enable educators and designers to superimpose virtual graphics over real objects. In addition, they allow users to interact with digital content via physical manipulation (Wei et al., 2015). AR facilitates more effective demonstrations of spatial and temporal concepts, as well as the contextual relationships between real and virtual objects. It has been shown to not only increase individuals’ learning motivation but also improve learning
performance (Zhang et al., 2014; Wei et al., 2015). However, many of the aspects that allow learners to interact with real objects when using AR systems are still to be discovered (Zhang et al., 2014; Chang et al., 2015b).

Although AR technologies have been around for more than 50 years, the recent proliferation of mobile devices has made affordable AR systems available to the general public (Sommerauer & Müller, 2014). AR systems have been widely applied in education (Sayed et al., 2011; Ibáñez et al., 2015; Wei et al., 2015) and increasingly applied with mobile devices (such as smartphones or tablets) to assist learning (Furió et al., 2013; Hsiao et al., 2016; Huang et al., 2016a). Mobile AR applications leverage the built-in cameras, GPS sensors, Internet access, and mobile devices to embed real world environments with dynamic, context-aware, and interactive digital contents (Sommerauer & Müller, 2014). To enhance user experience, AR technologies embed digital information on objects or places in the real-world (Zhang et al., 2014; Hsiao et al., 2016). AR provides great opportunities for online teaching that emphasizes the practical training and is particularly used in non-classroom training (Hsiao et al., 2016).

Prior studies have indicated that guidance and information provided by learning partners can improve learning performance (Webb, 1982). When people feel isolated in their learning process, their learning satisfaction is reduced (Hiltz & Wellman, 1997; Rovai & Wighting, 2005). Therefore, the interactive learning environment created by AR also enhances learning satisfaction among learners (Hsiao et al., 2016; Huang et al., 2016a). AR offers a multiplicity of applications within the field of education. To complement u-learning with real-time information, AR and RFID functions were applied to create enhanced outdoor u-learning environments for the study of natural science (Liu et al., 2009). Sayed et al. (2011) developed AR student cards that incorporated brand new data-processing and interactive modes to help students visualize their learning goals. Manipulative aids were proved as an effective learning tool for interactivity and usefulness of AR (Hsiao et al., 2016).

Cognitive style

Most innovative technologies for education ignore the fact that students with different levels of achievement require different designs of learning systems and learning strategies (Huang & Chiu, 2015). Thus, we further consider the individual difference such as cognitive style in our study. Cognitive style refers to an individual’s typical or habitual mode of thinking, perceiving, remembering, and problem solving (Soflano et al., 2015; Chang et al., 2015a). Cognitive style is defined as a kind of personal characteristic that processes and organizes information and experience based on personal preference (Messick, 1984). It has been considered as individual differences that can be used to affect the adoption of digital learning systems (Chang et al., 2015a). The personal differentiations of cognitive styles can be categorized into perception, thinking, learning, and problem solving (Witkin et al., 1975). Cognitive style is an approach that an individual habitually and preferentially uses to organize and present information. Individuals consistently adopt the same approach and these preferred modes of processing information differ among individuals (Witkin et al., 1975). Individuals with different cognitive styles use different approaches in learning as well. Prior studies have examined the impacts of cognitive styles on learning and found learners with different cognitive styles have different preferences (Chang et al., 2015a).

Prior studies have proved that matching cognitive styles with learning activities can lead to more efficient learning and better learning achievement (Ford & Chen, 2001). Thus, matching the cognitive styles of learners with the instruction approach contribute significantly to learning performance (Chang et al., 2015a). Because of the individual cognitive differences among learners, no one instructional method is appropriate for the array of cognitive styles. Researchers have been encouraging instructors to learn a diverse set of teaching skills so that they complement the array of cognitive styles of their students (Easton, 2003). Fitting instructional methods to learners’ cognitive styles not only improves their learning performance but also attitude toward learning (Dunn & Dunn, 1994). To increase the efficacy and satisfaction of learning experience, the design of an instruction system must consider learners’ cognitive styles (Hsieh et al., 2011). However, the performance designs that address diverse cognitive styles within new learning technologies remains inconclusive (Dillon & Gabbard, 1998). This knowledge gap leads us to examine the potential of an AR featured u-learning system with that integrates different cognitive styles.

There are different measures of cognitive styles. We choose “Field Dependence (FD)/Field Independence (FI) theory” (Witkin et al., 1975) among the cognitive styles because this theory fits our research context by using AR in the u-learning environment. “FD/FI theory” is the most common applied measure of cognitive styles. The word “field” can be a set of thoughts, ideas, or feelings (Witkin et al., 1975). FD learners refer to individuals that rely highly on structural format and prefer guidance in their learning process (Jonassen & Grabowski, 1993; Ford & Chen, 2000). In other words, FD learners are not good at constructing and analytical activities. The learning...
experience of FD individuals highly relied on the stimulus of external information and interactivity (Ford & Chen, 2000). On the contrary, FI learners prefer internal references instead of structured format and they are normally independent thinkers (Jonassen & Grabowski, 1993). That is, FI learners prefer analytical and active learning approaches (Frank & Keene, 1993). Field-mixed (FM) learners refer to individuals who do not have a clear FD or FI orientation but rather fall in the middle of the FD and FI spectrum (Liu & Reed, 1994). AR featured u-learning system intends to offer a real-life and interactive learning environment. In this case, FD/FI theory can highlight the different cognitive styles in explaining the learning performance in the AR featured u-learning environment.

**Research method**

This section shows our system design and development, experimental design, and selections of participants.

**System design and system development**

In recognition of Kolb’s (1984) idea that mature learning environments combine both learning and life, we developed a Ubiquitous Learning Instruction System with Augmented Reality Features (UL-IAR). The design incorporated AR technology and extensive computing power within a context-aware u-learning system to provide adaptive learning strategies to assist English learning. Figure 1 shows the system structure of UL-IAR.

UL-IAR is an English learning system which integrates AR and Wikitude World Browser (See Figure 2). This system was developed on the Android smartphone using Wikitude SDK and Android SDK. The Wikitude World Browser was a mobile software that integrates AR technologies, GPS, and mobile networks which allowed users to retrieve the information relevant to their surroundings (Wikitude, 2016). The primary features of UL-IAR include GPS positioning, the highlighting of local features, mark-up, scaffolding instruction, and real-time tests. Current applications of Wikitude World Browser were commonly employed in business contexts such as providing shopping information and popular scenic spots. However, applications within educational contexts
were rare. This UL-IAR system which integrated learning strategies and real-time quizzes helped users to learn English in real life contexts was one of the pioneering applications in the field of education.

![Figure 3. English Vocabulary in the smartphone](image)

![Figure 4. Real time exam](image)

![Figure 5. Result (Scores) of real-time examine](image)

The system structure of UL-IAR can be illustrated as following. First, users must turn on their GPS function in their mobile devices to increase the precision of geographic positioning (See Figure 3). After that, when users turned on UL-IAR, they can choose their current location to start learning English. UL-IAR connected the database through GPS or wireless Internet to provide scenery related instruction and attendant learning materials such as vocabulary or sentences that applicable in real-life contexts (See Figure 4). Third, users can evaluate their learning performance through real-time tests which were served as the pretest in this study. Based on the results of pre-tests, UL-IAR will provide different learning strategies (enforcing, semi-enforcing, and non-enforcing instructions) to assist users’ prior learning (See Figure 5). Last, after the scaffolding instructions, there will be another test to investigate users learning achievements when using different learning strategies.

**Experimental design**

There were two steps before the experiments. First, the participants were asked to take the “Embedded Figures Test (EFT)” to identify their cognitive styles (Messick, 1962). The EFT test included two parts with each part composed of 16 questions embedded with a simple figure among numerous complex figures (Messick, 1962). The completion time was limited to 20 minutes. Second, after identifying the cognitive styles, it took around 10 minutes to introduce and allow the participants to practice the UL-IAR system. Participants were asked to use the ULIAR system as the practice before visiting Kaohsiung West Bay to conduct the experiment. Overall, there were 30 minutes in the pre-experiments steps.

The EFT test was conducted with continuous scoring with a range from 0 (FD) to 32 (FI) to discern the cognitive styles among FD/FI users. This research categorized both the top and bottom 27% of users into both the FI and FD groups wherein the 46% of learners that fall between both ends were defined as the FM users (Spanjer &
Specifically, FI and FD individuals are those who are polarized at the end of the spectrum with extremely high and low scores respectively (Liu & Reed, 1994). We found that there were more FM participants followed by FD and FI in our sample size. It is easily to have more FM users in the sample size because 46% of the individuals that falls in the middle of the scores in EFT. The distribution is also verified in other studies (Angeli et al., 2009; Nicolaou & Xistouri, 2011). To have equal sample size for three different learning strategies, there were 30 FD, FI, and FM users selected for the experiment. Finally, we selected 90 participants to make sure there were equal number of participants represented in each of the three cognitive styles.

There were two steps in the experiment. First, the participants were asked to use UL-IAR in Kaohsiung West Bay to experience and learn English (See Figure 6). After that, they took the first test (pre-test) to exam what they have learned from UL-IAR. Second, different learning strategies were assigned to the participants to conduct the review of what they have learned from the first round. The participants used the UL-IAR system for the review of learning contents. In the end, there was a second test. Overall, the experiments were around 60 minutes.

We conducted field experiments to investigate users’ learning performance while using UL-IAR. Although field experiments can’t control all variables, the experiment results are closest to the real-life learning context. We intended to find out the most appropriate learning strategies for FD, FM, and FI individuals in using UL-IAR. We chose Kaohsiung West Bay as our experiment field because it has many popular scenic spots and numerous street food vendors. The West Bay scenic spots include National Sun Yat-sen University, West Bay Beach Hall, the British Consulate, No. 1 Ship Canal Landscape Bridge, West Bay Tunnel, and Gushan Ferry Station. Our learning materials and in-app tests were developed and based on these scenic spots and street foods which provided vivid learning contents. All of the learning materials and contents were revised by English native speakers.

Learning strategies are defined as the behaviors and thoughts that individual engages during learning and that are intended to influence his/her encoding process.” Enforcing learning strategy is considered as an appropriate learning strategy in our research context because it can assist individuals to accomplish their learning task more
effectively (Wood et al., 1976). Enforcing strategy is defined to help learners recapitulate and integrate what was explicitly displayed in the instructional materials and techniques that contribute to the memorization and understanding of learning contents. Enforcing strategy has been approved to be an effective learning strategy that leads to positive effects on individuals’ learning performance. We applied enforcing strategy with different levels in our experiment: enforcing, semi-enforcing, and non-enforcing to explore the learning effects of users with different cognitive styles. With enforcing strategy, the UL-IAR showed the vocabulary of the scenic spots that learners did not know in the first exam. This incorrect vocabulary was linked to the scenic spots with a green “plus” mark (See Figure 7). The enforcing strategy emphasizes the vocabularies that are incorrect in the first exam. With semi-enforcing strategy, the UL-IAR would show all the scenic spots. When learners did not know the vocabulary of scenic spots in the first exam, the UL-IAR would show a read “cross” mark (See Figure 8). If learners already knew the vocabulary of the scenic spots, UL-IAR system would not have any highlight. With non-enforcing strategy, no matter learners knew the vocabulary or not, there was no specific highlight in the UL-IAR (See Figure 9). It only showed the vocabulary that is related to the scenic spot.

Figure 9. Non-enforcing strategy

Experiment participants

We recruited and selected participants by using convenient and snowball sampling. To ensure the variety and heterogeneity of participants, participants were selected from different ages and backgrounds. Overall, there were 90 participants in our experiment who ranged from the age of 18 to 30 years old. They were college students, medical care workers, service industry employees, social workers, and kindergarten teachers. The objective of this experiment was to separate the appropriate learning strategies for FI, FD, and FM users when using our UL-IAR. We did not employ control and experimental groups but randomly assigned participants of the three different cognitive styles into three different learning strategies. Therefore, there were 9 groups which contain 10 participants each. To measure group differences, cell size with 30 can explain for 80% power. If decreased, no lower than 7 per cell (VannVoorhis et al., 2007). We have 10 participants in each cell which is acceptable sample size.

Research results and discussion

This research aimed to investigate the learning performance of using UL-IAR system among users that have different cognitive styles (field dependent (FD), mixed field (FM), and field independent (FI)). The UL-IAR system embodied different learning strategies (enforcing, semi-enforcing, and non-enforcing) to investigate their impact on users’ learning performance. Thus, the independent variables were cognitive styles and learning strategies. The dependent variable was the learning performance of using UL-IAR to learn English. We first conducted a descriptive statistical analysis of the learning performance of all users (See Table 1). The results showed that the average learning scores of FD users with enforcing strategies were the highest (Mean = 91.25). This implies that users with FD cognitive styles matched the enforcing learning strategies. Second, individuals with FM cognitive styles had the highest scores with enforcing learning strategy (Mean = 81.625). Third, the average scores of FI users with semi-enforcing learning strategy was the highest one among the other three. This implies FI users preferred a semi-enforcing learning strategy which would lead to greater learning performance.

We conducted two-way ANCOVA analysis (See Table 2) and the results showed that cognitive styles were not statistically significant ($F = 1.995, p = .143$). However, learning strategies were statistically significant ($F = 10.426; p = .000$). In addition, we found that there were interactive effects between cognitive styles and learning strategies ($F = 3.493; p = .011$) (See Table 2). The partial Eta Squared value indicates the effect size and should be compared with Cohen’s guidelines (0.2 - small effect, 0.5 - moderate effect, 0.8 - large effect). It can be seen that effect size of the interaction between Cognitive Styles and Learning Strategies is small (0.149). This value describes how much of variance in the dependent variable is explained by the independent variable (14.9%).
This implied that the impact of cognitive styles on learning performance is affected by another independent variable (learning strategies). A statistical interaction occurs when the effect of one independent variable (cognitive styles) changed depending on the level of another independent variable (learning strategies). Thus, we need to examine how learning strategies are affected by cognitive styles. We further conducted the simple effect tests (Weinberg & Abramowitz, 2002).

Table 1. The descriptive statistics of the learning effect

<table>
<thead>
<tr>
<th>Cognitive style</th>
<th>Learning strategy</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Total SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD</td>
<td>Enforcing</td>
<td>91.250</td>
<td>5.270</td>
<td>10</td>
<td>16.563</td>
</tr>
<tr>
<td></td>
<td>Semi-enforcing</td>
<td>81.875</td>
<td>10.396</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-enforcing</td>
<td>63.3725</td>
<td>17.348</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>FM</td>
<td>Enforcing</td>
<td>81.875</td>
<td>5.472</td>
<td>10</td>
<td>15.110</td>
</tr>
<tr>
<td></td>
<td>Semi-enforcing</td>
<td>80.6250</td>
<td>16.783</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-enforcing</td>
<td>71.8750</td>
<td>18.923</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>FI</td>
<td>Enforcing</td>
<td>76.7500</td>
<td>18.049</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Semi-enforcing</td>
<td>77.4500</td>
<td>13.066</td>
<td>10</td>
<td>14.252</td>
</tr>
<tr>
<td></td>
<td>Non-enforcing</td>
<td>73.0750</td>
<td>12.072</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Enforcing</td>
<td>83.292</td>
<td>12.502</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Semi-enforcing</td>
<td>79.983</td>
<td>13.324</td>
<td>30</td>
<td>15.222</td>
</tr>
<tr>
<td></td>
<td>Non-enforcing</td>
<td>69.440</td>
<td>16.403</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

Note. FI = field-independent, FM = mixed-field, FD = field-dependent.

Table 2. Two-way ANCOVA analysis of the learning effect

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.*</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>4929.881</td>
<td>1</td>
<td>4929.881</td>
<td>36.488</td>
<td>.000</td>
<td>.313</td>
</tr>
<tr>
<td>Cognitive Styles</td>
<td>538.954</td>
<td>2</td>
<td>269.477</td>
<td>1.995</td>
<td>.143</td>
<td>.047</td>
</tr>
<tr>
<td>Learning Strategies</td>
<td>2817.209</td>
<td>2</td>
<td>1408.605</td>
<td>10.426</td>
<td>.000***</td>
<td>.207</td>
</tr>
<tr>
<td>Cognitive Styles x Learning Strategies</td>
<td>1887.817</td>
<td>4</td>
<td>471.954</td>
<td>3.493</td>
<td>.011*</td>
<td>.149</td>
</tr>
<tr>
<td>Error</td>
<td>10808.739</td>
<td>80</td>
<td>135.109</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

The results of the simple effect tests show that an enforcing learning strategy influenced three different cognitive styles users (FD, FI, and FM) \( (F = 5.996**) \) (See Table 3). Since we have 3 groups of participants in three cognitive styles, the significant levels are divided by 3. It implies that the enforcing learning strategy differs in the learning performance among three different cognitive styles of UL-IAR users. On the contrary, semi-enforcing \( (F = 0.500) \) and non-enforcing \( (F = 2.483) \) learning strategies did not show statistically significant difference.

Table 3. The simple effect tests of learning strategies to cognitive styles

<table>
<thead>
<tr>
<th>Learning strategies</th>
<th>FD</th>
<th>FI</th>
<th>FM</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean / SD</td>
<td>Mean / SD</td>
<td>Mean / SD</td>
<td></td>
</tr>
<tr>
<td>Enforcing</td>
<td>91.250 / 5.270</td>
<td>76.750 / 18.049</td>
<td>81.875 / 5.472</td>
<td>5.996**</td>
</tr>
<tr>
<td>Semi-enforcing</td>
<td>81.875 / 10.396</td>
<td>77.450 / 13.066</td>
<td>80.625 / 16.783</td>
<td>0.500</td>
</tr>
<tr>
<td>Non-enforcing</td>
<td>63.373 / 17.348</td>
<td>73.075 / 12.072</td>
<td>71.875 / 18.923</td>
<td>2.483</td>
</tr>
</tbody>
</table>

Note. FI = field-independent, FM = mixed-field, FD = field-dependent. *p < .05/3; **p < .01/3.

Since enforcing learning strategy differs in the users’ learning performance among three cognitive styles, we further examine the simple effect tests for enforcing learning strategies. Among three cognitive styles (FD, FI, and FM users), FD users aligned with an enforcing learning strategy yield the best English learning performance \( (p = .002) \) (See Table 4). With enforcing learning strategy, the learning performance of FD users is better than that of FI users \( (\text{mean difference} = 16.617, p = .002**) \).

Alternatively, we examined whether different cognitive styles with different learning strategies lead to different learning performance (See Table 5). The results showed that FD users’ learning performance was significantly different with three learning strategies (enforcing, semi-enforcing, and non-enforcing) \( (F = 21.894**) \). On the contrary, FI and FM users’ learning performance did not show significantly different with three different strategies \( (F = 0.764 \text{ and FM; } F = 1.049) \).
Table 4. The simple effect tests for enforcing learning strategy among three cognitive styles

<table>
<thead>
<tr>
<th>(I) Cognitive Style</th>
<th>(J) Cognitive Style</th>
<th>Mean Difference (I-J) / SE</th>
<th>Sig.</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD at Enforcing</td>
<td>FI at Enforcing</td>
<td>16.617/4.876</td>
<td>.002**</td>
<td>FD &gt; FI</td>
</tr>
<tr>
<td>FM at Enforcing</td>
<td>FI at Enforcing</td>
<td>11.069/4.838</td>
<td>.031</td>
<td></td>
</tr>
<tr>
<td>FI at Enforcing</td>
<td>FD at Enforcing</td>
<td>-16.617/4.876</td>
<td>.002**</td>
<td></td>
</tr>
<tr>
<td>FM at Enforcing</td>
<td>FD at Enforcing</td>
<td>-5.548/4.773</td>
<td>.256</td>
<td></td>
</tr>
<tr>
<td>FI at Enforcing</td>
<td>FD at Enforcing</td>
<td>-11.069/4.838</td>
<td>.031</td>
<td></td>
</tr>
<tr>
<td>FI at Enforcing</td>
<td>UL at Enforcing</td>
<td>5.548/4.773</td>
<td>.256</td>
<td></td>
</tr>
</tbody>
</table>

Note. FI = field-independent, FM = mixed-field, FD = field-dependent. *p < .05/3; **p < .01/3.

Table 5. The simple effect tests of cognitive styles to learning strategies

<table>
<thead>
<tr>
<th>Cognitive styles</th>
<th>Enforcing Mean / SD</th>
<th>Semi-enforcing Mean / SD</th>
<th>Non-enforcing Mean / SD</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD</td>
<td>91.2500 / 5.270</td>
<td>81.875 / 10.396</td>
<td>63.373 / 17.348</td>
<td>21.894**</td>
</tr>
<tr>
<td>FI</td>
<td>76.750 / 18.049</td>
<td>77.450 / 13.066</td>
<td>73.075 / 12.072</td>
<td>0.764</td>
</tr>
<tr>
<td>FM</td>
<td>81.875 / 5.472</td>
<td>80.625 / 16.783</td>
<td>71.875 / 18.923</td>
<td>1.049</td>
</tr>
</tbody>
</table>

Note. FI = field-independent, FM = mixed-field, FD = field-dependent. *p < .05/3; **p < .01/3.

To understand the interactive effect between cognitive styles and learning strategies, we conducted the simple effect tests of cognitive styles to learning strategies as the post-hoc analysis (See Table 6). Among three learning strategies, FD users with enforcing learning strategy yield best English learning performance which was better than semi-enforcing learning strategy (Mean = 81.875; p = .007) and non-enforcing learning strategy (Mean = 63.373; p = .000). FD users with semi-enforcing learning strategy was better than FD users with non-enforcing (p = .001). Overall, when users’ cognitive style was FD, enforcing learning strategy was better than semi-enforcing and non-enforcing learning strategy.

Table 6. The simple effect tests for FD users among learning strategies

<table>
<thead>
<tr>
<th>(I) Learning strategies</th>
<th>(J) Learning strategies</th>
<th>Mean difference (I-J) / SE</th>
<th>Sig.</th>
<th>Note</th>
</tr>
</thead>
</table>

Note. FI = field-independent, FM = mixed-field, FD = field-dependent. *p < .05/3; **p < .01/3; ***p < .001/3.

Conclusion

Researchers have been interested in promoting context-aware u-learning to improve learning performance. How to better apply new technologies to create vivid learning environment and experience is a critical research issue. This paper has two main objectives. The first objective of this study is to develop a u-learning system that incorporates AR technology (UL-IAR) to improve English learning performance and to animate English learning experience. Based on this UL-IAR system, we further developed English learning content with different learning strategies (enforcing, semi-enforcing, and non-enforcing). The second objective is to examine whether cognitive styles and learning strategies affect learning performance when using UL-IAR. We further tested the optimum alignment of learning strategy and cognitive style for the UL-IAR application. Our main findings can be summarized as following. First, different learning strategies with different cognitive styles lead to different learning performance when they use UL-IAR to learn English in real life contexts. Second, field dependent users aligned with an enforcing learning strategy yield the best English learning performance by using UL-IAR. Applying an enforcing learning strategy for FD users in UL-IAR is more productive than applying either a semi-enforcing or a non-enforcing learning strategy. Third, FD users benefit more from an enforcing learning strategy than users who have FI or FM cognitive styles. The results show that using the UL-IAR with an enforcing learning strategy can positively improve FD users’ learning performance. Our findings are consistent with Dunn and Dunn’s (1994) argument that an individual’s learning performance or learning attitude can be positively improved when teaching and learning resources match an individual’s cognitive style.

Overall, our contributions can be summarized as following. First, we developed a context-aware u-learning system with augmented reality technology (UL-IAR) to allow users to animate their English learning experience.
in a real life context. UL-IAR allows users to listen, to speak, and experience English learning in the real life context. This study provides practitioners a reference to develop and implement learning system in education field. Second, the UL-IAR allows individuals to learn English in a real life context which improves the understanding of English vocabulary and enlivens learning experience. This study echoes Kolb’s idea that mature learning environments combine both learning and life (Kolb, 1984). Third, applying learning strategies to users with different cognitive styles leads to different learning performance in using UL-IAR. Applying enforcing learning strategy in UL-IAR to FL users leads to best learning performance. This finding suggests the best alignment of learning strategies with users of different cognitive styles. Forth, if we apply learning strategies in the learning processes by using UL-IAR, users immediately recognize areas of difficulty and are able to address them when they arise. This finding implies that applying learning strategy in UL-IAR system can contribute to better learning performance. Our findings also echo prior research that suggests that context-aware u-learning can improve individuals’ learning performance (Sandberg et al., 2011, Cheng et al., 2010).

There are a few limitations in this research. First, we didn’t find adaptive relationships in some learning strategies and cognitive styles. Future research may test different learning strategies for different cognitive styles to find out better alignments for different users. Second, as AR technologies mature, to create authentic language u-learning environment with AR is emerging as a promising approach for improving learning performance. This study demonstrates such design and finds the relationships between learners’ cognitive styles, learning strategies and learning performance in the UL-IAR usage. Future research can provide a broader and deeper view of how different learners can be adaptively supported in using UL-IAR to learn English effectively.

Acknowledgements

The authors thank the partial support from Ministry of Science and Technology, Taiwan, under research grant numbers NSC102-2410-H-017-020-MY2, MOST 105-2410-H-017-004-, MOST104-2511-S-269-002-MY3, and MOST104-2511-S-110-009-MY3.

References


The Effects of Using the Kinect Motion-sensing Interactive System to Enhance English Learning for Elementary Students

Wen Fu Pan
Department of Educational Administration and Management, National Dong-Hwa University, Taiwan // s1210@gms.ndhu.edu.tw

ABSTRACT
The objective of this study was to test whether the Kinect motion-sensing interactive system (KMIS) enhanced students' English vocabulary learning, while also comparing the system's effectiveness against a traditional computer-mouse interface. Both interfaces utilized an interactive game with a questioning strategy. One-hundred and twenty students were chosen for the experiment. The students were divided into two groups: Kinect, computer-mouse, and control. The participants' vocabularies were evaluated three times during a pre-test, a post-test, and a 1-month post-test. The following results were obtained: (1) there was a partially disordinal interaction relationship between the three groups and the three tests. Post-hoc comparison showed that the three tests have an order relationship. (2) The within group comparisons, both for the motion-sensing and computer-mouse groups which utilized an interactive game with a questioning strategy, displayed a relatively significant long-term retention. (3) In the between group comparison, the two interactive groups (computer-mouse and motion-sensing group) did not reach significant difference in English vocabulary learning. This means the motion-sensing interface of the KMIS was not a key-factor to affecting short-term or long-term learning retention. Therefore, our suggestion is that teachers can adopt interactive games with a questioning strategy to enhance students' long-term English vocabulary retention.

Keywords
Computer assisted learning, Human-computer interaction, Kinect sensor, Motion-sensing systems, Vocabulary learning

Introduction
Traditional one-way rote memorization method for learning English vocabulary is frequently found in schools (Smith, Li, Drobisz, Park, Kim, & Smith, 2013). However, the effects of such methods have not been found to be better than interaction (Ge, 2015). Vygotsky (1978) considered that in second language acquisition, learners need to interact with the socio-cultural environment via artifacts. These artifacts are referred to as “interfaces” between the subject and object from the viewpoint of human-computer interaction (Engeström, 2000).

Early human-computer interface (HCI) studies mostly adopted usability testing (Buur & Bedker, 2000). In the 1990s, some scholars began to cite activity theory, proposed by Leont’ev in the 1930s, as a theoretical framework of HCI design (Kaptelinin, 1996; Kuutti, 1996; Nardi, 1996). Activity theory emphasizes how to construct meaning from interaction between subject and object via artifacts (such as rules, books, etc.) (Leont’ev, 1974). Subsequently, activity theory also became one of the theoretical frameworks for language learning (Oxford, 1990). In 2003, Bedny and Karwowski divided activities into the following five levels: activity, task, action, operation, and function; they also incorporated two design types: subject-oriented and object-oriented, based on their proposed Systemic-Structural Theory of Activity (Bedny & Harris, 2005).

Subject-oriented design focuses on a subject’s socio-cultural context and has often been adopted by studies of second language learning (Chapelle, 2009). On the other hand, in order to assess the usability of an emerging technology, researchers have often adopted object-oriented design (Munassar & Govardhan, 2011). This study also adopts a type of object-oriented design called “object-mental action” from activity theory. Specifically, a subject (the learner) interacts with an object (game-based animation) via the Kinect Motion-sensing Interactive System.

Edgar Dale’s cone of experience theory indicates that two-way interactive learning helps learners to obtain up to 90% learning retention (Dale, 1969). Human-computer interaction also benefits learning retention (Papastergiou, 2009; Prensky, 2005); however, is this effect derived from the human-computer “interactive content,” or “operating interface”? This question is worthy of further research. Therefore, in this study, we designed a game-based learning activity as the interactive “content” and a motion-sensing operation as the interactive “interface” for English vocabulary learning. The related literature is reviewed as follows.
Applying game-based learning with a questioning strategy as interactive content

Previous research has shown that game-based English learning has resulted in better retention than traditional rote memorization (Flores, 2015). Hwang, Chiu, and Chen (2015) also indicated that game-based learning is able to improve students’ inquiry-based learning performance, especially in an interactive environment. Also, enjoying the game was cited as an important reason why students were willing to finish interactive tasks (Star, Chen, & Dede, 2015). The design of digital games is an important and often used method for enhancing learning motivation. A learner’s motivation to participate is enhanced through game-based learning (Birk, Atkins, Bowey, & Mandryk, 2016; Ronimus & Lyytinen, 2015). The goal of this study is to design an English vocabulary learning activity that integrates digitized game-based interaction.

In addition, a questioning strategy was implemented in this study to enhance the two-way interactive learning. The questioning strategy is defined as actively presenting a question and waiting for the students’ answer. Research has indicated that implementing a questioning strategy in English learning can also result in better retention (Basturkmen, 2001; Boyd & Rubin, 2006; Shomoossi, 2004; Yang, 2010).

Applying motion-sensing operation as an interactive interface

Developments in emerging technology have transformed the types of interaction between humans and computers. Past research (Chuang & Kuo, 2016; Hsiao & Chen, 2016; Sheehan & Katz, 2012) has shown that applying various motion-sensing technology to learning environments benefits students. Microsoft released the source code of Kinect (3D depth sensor) in 2012, and since then there has been a lot of development in motion-sensing applications. Currently all short-distance motion (or gesture) sensors use either infrared light emitters and sensors, ultrasonic sensors, or 3D depth sensors (Kinect) (Kumaragurubaran, 2011). The principle of the Kinect motion-sensing technique is that it employs three lenses, as well as a diffuser lens to expand or diffuse projected laser speckles. For the speckles that reach the human body, a separate camera coordinated with a light coding technique is employed to collect the 3D depth of field information regarding the human body within a 5-m tapered space (Pan, Chien, & Tu, 2012a). Pan et al. (2012a) compared the differences between infrared, ultrasonic, and Kinect sensing techniques. Applying Kinect in learning offers the following benefits: (1) it does not require a handheld controller; (2) it provides real-time feedback; (3) it is able to distinguish humans from objects; (4) it provides teachers (or developers) with a way to customize interactive content.

A lot of research has been conducted and is ongoing in applying Kinect technology to various fields (Nissimov, Goldberger, & Alchanatis, 2015; Yao, Wang, Cai, & Zhang, 2015). Kinect motion-sensors have been integrated into interactive learning, and research in this area has become a growing trend (Chuang & Kuo, 2016). For example, Sommool, Battulga, Shih, and Hwang (2013) applied Kinect motion-sensing technology to create and evaluate interactive learning classrooms; Tutwiler, Lin, and Chang (2013) applied it to multiple intelligence instruction; and Levinger, Zeina, Teshome, Skinner, Begg, and Abbott (2016) utilized Kinect in gait practice for knee replacement rehabilitation. In recent years, Pan led a team focused on the development of Kinect applications for educational situations (Pan, Tu, & Chien, 2014). Their research covered a range of applications including campus safety (Pan, Chien, Liu, & Chan, 2012b), accessible learning (Pan et al., 2012a), and interactive learning (Pan, Lin, & Wu, 2011). Pan (2013) indicated that the Kinect motion-sensing interface can better enhance students’ learning motivation compared to a more traditional computer-mouse interface. Some studies (Sommool et al., 2013; Vrellis, Moutsioulis, & Mikropoulos, 2014; Yuan, Hsieh, Chew, & Chen, 2015) have also supported the idea that the novelty of the Kinect system can attract students’ attention and increase learning motivation.

Concept framework of this study

Based on the above literature review, the concept framework of this study is shown in Figure 1.

A one-way interaction rote memorization method is frequently used in schools for learning of English vocabulary (Smith et al., 2013). In order to improve learning methods, two-way interactive learning and human-computer interfaces should be applied based on activity theory. The Kinect Motion-sensing Interactive System, or KMIS, designed in this study includes two parts: (1) applying game-based learning with a questioning strategy as interactive content; (2) applying the Kinect motion-sensing operation as an interactive interface. Past research (Pan, 2013; Sommool et al., 2013; Vrellis et al., 2014; Yuan et al., 2015) regarding Kinect motion-sensing applications in learning has usually adopted an empirical method, so an experimental design was used in

189
this study. In addition, although some studies (Pan, 2013; Vrellis et al., 2014; Yuan et al., 2015) have also made the comparisons between Kinect motion-sensing and computer-mouse interface, these studies did not focus on the interactive relationship of the learning “content” and “interface,” or its effects on English learning. Therefore, through this experimental design, the researcher wants to assess the within-group (interactive content) and between-group (operating interface) interactive relationship, as well as compare their effects on short-term and long-term English vocabulary retention. It was expected that the findings from this KMIS design would show improvement over one-way rote memorization learning.

Objectives of the study

According to the above analysis, this study has the following three research purposes:

- To do statistical analyses on the interactive relationship between three groups (no-interaction, Kinect motion-sensing and computer-mouse) and three tests (pre-test, post-test and delayed post-test).
- To analyze the performances on the three tests (pre-test, post-test and delayed post-test) in interactive learning with an integrated questioning strategy game for English vocabulary learning.
- To compare the short-term and long-term retention effects on English vocabulary learning in the three groups.

Methods

Experimental design

The quasi-experimental design was adopted as shown in Table 1. The subjects of the experiment were divided into three groups: motion-sensing, computer-mouse, and the control group (no interaction). The motion-sensing group ($X_1$) was tested with the KMIS and the computer-mouse group ($X_2$) was tested with a traditional
computer mouse interface. Both the motion-sensing group ($X_1$) and computer-mouse group ($X_2$) had the same interactive content (a football game with a vocabulary quiz). Each group contained forty 6th grade students, who each completed the English vocabulary test three times: a pre-test ($O_1$), a post-test ($O_2$), and a 1-month delayed post-test ($O_3$). Types of interaction (three groups) and the three tests were the two factors in data analyses for a two-way mixed-design ANOVA. The two-way mixed-designed ANOVA was a better choice than analysis of covariance (ANCOVA) in this case because it was not only able to analyze the interactive relationships of three groups and three tests, but also able to compare the within-group and between-group differences.

<table>
<thead>
<tr>
<th>Table 1. Two-way mixed-design structure ($N=120$) with distinct types of interaction and tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment group</strong></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Control group</td>
</tr>
<tr>
<td>Computer-mouse group</td>
</tr>
<tr>
<td>Motion-sensing group</td>
</tr>
</tbody>
</table>

Note. $O$ represents a 25 multiple-choice question test; the motion-sensing group ($X_2$) used the Kinect motion-sensing interactive system (KMIS) as the interactive interface.

Research subjects

Six classes were randomly selected as research subjects with cluster sampling from 6th grade classes at a large-scale elementary school in Hualien City, Taiwan. Then two classes were randomly assigned to each of the three groups. Each group had 40 students averaging 12 years old, and the three groups had a total of 120 participants. Each group's participants had similar academic performance in school, and the subjects of the three groups were examined by the homogeneity test. Statistically, the Box's test did not reach the level of significance ($F = 1.522, p = .108$) and the $F$-test for the pre-test did not reach the level of significance for difference ($F = .22, p = .80$, shown in table 6). Therefore, the basic background of subjects and environmental factors of the three groups were considered homogenous, and the two-way mixed-designed ANOVA was adopted. The experiment and test data were collected from October to December of 2013.

Development of the English vocabulary cognition test

The English Vocabulary Cognition Test (EVCT) for 6th grade students was created for evaluating students' learning performance. To establish the content of the test, forty words were randomly sampled from the "1200 English Vocabulary Words" endorsed by the Ministry of Education in Taiwan for elementary and junior high school students. These words were drafted into 40 multiple-choice questions for a test for 6th grade students in another elementary school in Hualien City, Taiwan. The 169 valid tests were ordered by scores, and the top and bottom 27% of the tests were grouped into high-score and low-score for item analysis of the questions. Under the significant threshold of $p < .013$, twenty-five of the best (excellent) questions were selected for the formal test. Table 2 shows the Item Number, Initial Item Number in the pilot test, Item Difficulty ($P$), Item Discrimination ($D$), Critical Ration ($CR$), and Significance ($p$). The $P$-values appeared in between .272 and .554 ($P$-values = (Percentage correct $High-score-group$ + Percentage correct $Low-score-group$) / 2; the excellent questions' $P$-values ideally should be between .2 and .8), and the $D$-values fell between .239 and .652 ($D$-values = $P_H - P_L$; all questions' $D$-values ideally should be at least .2 or more). The mean $P$-values of the 25 questions was .407 (for excellent questions the mean of $P$-value ideally should be close to .5), the mean $D$-values was .403 (for excellent questions the mean of $D$-values ideally should be close to 1.0), and the mean $CR$-values was 4.455 (all questions reached statistical significance, $p < .05$).

<table>
<thead>
<tr>
<th>Item no.</th>
<th>Item no. in pilot test</th>
<th>Difficulty ($P$)</th>
<th>Discrimination ($D$)</th>
<th>Critical ration ($CR$)</th>
<th>Significance ($p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>.478</td>
<td>.391</td>
<td>4.038</td>
<td><strong>.000</strong></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>.380</td>
<td>.457</td>
<td>5.054</td>
<td><strong>.000</strong></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>.380</td>
<td>.587</td>
<td>7.199</td>
<td><strong>.000</strong></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>.402</td>
<td>.543</td>
<td>6.316</td>
<td><strong>.000</strong></td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>.500</td>
<td>.565</td>
<td>6.500</td>
<td><strong>.000</strong></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>.467</td>
<td>.413</td>
<td>4.314</td>
<td><strong>.000</strong></td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>.500</td>
<td>.652</td>
<td>8.162</td>
<td><strong>.000</strong></td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>.522</td>
<td>.478</td>
<td>5.173</td>
<td><strong>.000</strong></td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>.359</td>
<td>.283</td>
<td>2.925</td>
<td>.004**</td>
</tr>
</tbody>
</table>
After item analysis, the final draft of the EVCT contained 25 multiple-choice questions composed of 10 word meaning questions, 9 word form questions, and 6 word usage questions. Each question provided 4 choices, with only one correct answer. Each correct answer was scored as 4 points for a full score of 100. The EVCT was used for the pre-test, post-test, and 1-month delayed post-test. It was also used for the content of the interactive learning game for both the Kinect and computer-mouse interface.

**Experimental process**

The experiment was performed from October to December 2013. The treatments of the three groups for their review activity phase are outlined in Table 3. The pre-test, post-test, and delayed post-test were arranged on October 17, November 7, and December 12 of 2013. Each group received identical content both for their EVCT and review activities. The control group merely viewed the test paper with the correct answers for their review activity while the motion-sensing and computer-mouse groups reviewed using the interactive football game on their respective interfaces. The only difference between the latter two groups was the interface, point and click vs. a physical kicking motion.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Pre-test</th>
<th>Review activities</th>
<th>Post-test</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 min</td>
<td>40 min</td>
<td>20 min</td>
<td>20 min</td>
</tr>
<tr>
<td>A1 Control group</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The volunteers viewed and practiced the test paper with answers.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2 Computer-mouse group</td>
<td>☑</td>
<td>The volunteers operated the game with computer mice, and the bystanders participated in watching and answering the vocabulary questions.</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>A3 Motion-sensing group</td>
<td>☑</td>
<td>The volunteers played the game with kicking motions, and the bystanders participated in watching and answering the vocabulary questions.</td>
<td>☑</td>
<td>☑</td>
</tr>
</tbody>
</table>

**Interactive system-design for the motion-sensing group**

The subjects were divided into the following three groups for the experiment: the control group, which did not have an interactive review activity; the computer-mouse group (X\(_1\)), which utilized a traditional point and click computer interaction; and the motion-sensing group (X\(_2\)), which utilized the Kinect motion-sensing interactive system (KMIS). The structure of the KMIS is described below.

The KMIS is a system composed of both software and hardware. Figure 2 shows the Kinect Software Development Kit (SDK) from Microsoft and the Kinect Flexible Action and Articulated Skeleton Toolkit.
(FAAST) offered by Suma, Krum, Lange, Koenig, Rizzo, and Bolas (2013) integrated with Windows 7, a projector, and Kinect hardware. When the participant stands at a proper distance from the Kinect sensor (about 1.5m-3.5m), the Kinect would convert the kicking motion into a computer command through FAAST and become a selection operation. The four footballs on the screen represent the four options for the multiple-choice questions. The researcher defined four kicking motions to respond to the four options: (1) the left foot kicking towards the left for the first ball (Option 1), (2) the left foot kicking forward for the second ball (Option 2), (3) the right foot kicking forward for the third ball (Option 3), and (4) the right foot kicking towards the right for the fourth ball (Option 4).

![Figure 2. The KMIS interactive system design for motion-sensing group](image)

**Integrated design of game-based interactive learning and a questioning strategy**

For the purposes of this research, the definition of “English vocabulary learning experience” is that learners read the vocabulary question on the review paper or screen, and then passively or actively find the correct answer from the review paper or game screen. The effects of their English vocabulary learning experience were evaluated by the three EVCT tests. The definition of “the questioning strategy” is that the questions of the English vocabulary game are actively shown on a projector screen, the learner’s answer is waited for, and finally a response is given by the learner (shown in Figure 3). The questioning strategy was only utilized with the computer-mouse group and motion-sensing group.

The computer-mouse group and the motion-sensing group respectively represented traditional and novel human-computer interaction. Both utilized the football game and questioning strategy (Figure 3). In contrast, the control group only reviewed the vocabulary test paper. The football game consisted of 25 multiple-choice questions where the participants gained points by selecting the correct answers. Whether or not the correct answer was selected, the screen would still display the correct answer as feedback. This feedback was an integral part of the questioning strategy which separated the subjects in the interaction groups (computer-mouse group and motion-sensing group) with the subjects in the control group, who only reviewed the paper tests with correct answers.

![Figure 3. Football game and questioning strategy integrated design for mouse and motion-sensing group](image)
The Flash football game used in this study was free software downloaded from the internet, at http://goo.gl/dFXLKw. The software was revised by K. H. Yen, a teacher at Li Xing Primary School (Yen, 2012). It allows users (teachers) to input customized questions and implement the questioning strategy to design learning activities with its football game.

**Data analysis method**

A two-way mixed-design ANOVA was utilized for testing the relationship between the three types of interaction (control or no-interaction, computer-mouse, motion-sensing) and the performance on the three tests (pre-test, post-test, delayed post-test). A homogeneity test for variance was performed before the analysis, and the data analyses were presented with a descriptive statistics summary, a two-way mixed-design ANOVA, a test of simple main effects, and the LSD Method.

**Results**

The homogeneity test for variance was performed before the two-way mixed-design ANOVA, and the result did not reach the level of significance (Box’s M = 18.972, F = 1.522, p = .108). This shows that the variance of the test scores was homogenous and that we could proceed with the successive statistical analyses. The results of the types of interaction and of the three tests, including a descriptive statistics summary, two-way mixed-design ANOVA, and a test of simple main effects, are shown in Table 4 to Table 7.

**The findings regarding descriptive statistics of the types of interaction (three groups) and the three tests**

Table 4 presents the population distribution, mean, and standard deviation of the types of interaction and the three tests. The overall mean of the three tests was also calculated for the post-test (M = 66.40, SD = 22.62), delayed post-test (M = 63.00, SD = 21.94), and pre-test (M = 55.33, SD = 22.06). Overall, the performance of all three groups improved, where the control group received the highest average score in the pre-test (M = 57.20), while the motion-sensing group received the highest average scores in the post-test (M = 71.30) and the delayed post-test (M = 65.60). The findings seem to reveal some variation within-group (between the three tests) or between-group. This result needs to be further tested with a mixed-design two-factor ANOVA.

**Table 4. Statistics summary of types of interaction (A) and three tests (B)**

<table>
<thead>
<tr>
<th>B. Three tests</th>
<th>A. Types of interaction</th>
<th>M</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1. Pre-test</td>
<td>A1. Control</td>
<td>57.20</td>
<td>22.16</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>A2. Computer-mouse</td>
<td>54.70</td>
<td>22.65</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>A3. Motion-sensing</td>
<td>54.10</td>
<td>21.82</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>55.33</td>
<td>22.06</td>
<td>120</td>
</tr>
<tr>
<td>B2. Post-test</td>
<td>A1. Control</td>
<td>60.70</td>
<td>22.22</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>A2. Computer-mouse</td>
<td>67.20</td>
<td>22.52</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>A3. Motion-sensing</td>
<td>71.30</td>
<td>22.41</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>66.40</td>
<td>22.62</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>A2. Computer-mouse</td>
<td>63.70</td>
<td>21.97</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>A3. Motion-sensing</td>
<td>65.60</td>
<td>22.17</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>63.00</td>
<td>21.94</td>
<td>120</td>
</tr>
</tbody>
</table>

**The findings of mixed-design two-factor ANOVA for the types of interaction and the three tests**

The two-factor analysis of types of interaction and the three tests is shown in Table 5. There was interaction (F = 14.98, p < .01, η² = .204) between the two factors (A×B), and the three tests (B) also showed significant variation (F = 114.62, p < .01, η² = .495). A partially disordinal interaction relationship is shown in Figure 4, where the students’ three tests before the interaction were initially ranked control group, computer-mouse group, and motion-sensing group, but the ranking was reversed in the post-test and delayed post-test. This change presents a significant interaction worth further analyses in the simple main effect test.
The findings of the simple main effect for the types of interaction and the three tests

The simple main effect is shown in Table 6. The three types of interaction (A) did not reach a significant level of difference ($F_{B1} = .22$; $F_{B2} = 2.28$; $F_{B3} = .75$) in the detailed items (B1.pre-test, B2.post-test, and B3.delayed post-test) of the three tests (B), revealing no between-group difference among the three types. In other words, the motion-sensing group was not superior to the computer-mouse and control groups in students’ vocabulary learning. The three tests (B), on the other hand, achieved the significant difference ($F_{A1} = 3.87$, $p < .05$, $\eta^2 = .090$; $F_{A2} = 65.00$, $p < .01$, $\eta^2 = .625$; $F_{A3} = 73.55$, $p < .01$, $\eta^2 = .653$) in the detailed items (A1.control, A2.computer-mouse, and A3.motion-sensing) of the types of interaction (A), presenting the necessity of post-hoc comparison of the three tests (B).

The findings of post-hoc comparisons for the types of interactive and the three tests

The post-hoc comparison (LSD Method) findings (Table 7) showed that the within-group (three tests) comparison has an order relationship (B2 > B1, $p < .05$). It revealed that the post-test performance was superior to the pre-test regardless of the group. The researcher considered that the practice effect may have affected short-
term learning retention. To further analyze each type of interaction, the students in the control group (A1) presented significantly higher post-test ($M_{A1} = 60.70$) than pre-test ($M_{B1} = 57.20$) performance ($B_2 > B_1, p < .05$). This reveals that simply memorizing vocabulary still presented a short-term memory benefit. However, there was no significant difference in the performance between delayed post-test and pre-test, implying that over a prolonged period of time the students forgot what they had memorized and statistical significance could not be achieved. The next group, the students in the computer-mouse group (A2) displayed an even more significant performance on the post-test ($M_{A2} = 67.20$) and delayed post-test ($M_{B2} = 63.70$) when compared to the pre-test ($M_{B1} = 54.70$) ($B_2 > B_1$ and $B_3 > B_1, p < .05$). This reveals that the traditional interaction of the computer-mouse group still led to a vocabulary learning effect. Finally, the learning performance of the motion-sensing group (A3) was analyzed. From Table 7, the LSD method result of the motion-sensing group was identical to that of the mouse group (post-test and delayed post-test performance were higher than pre-test), revealing that the type of interaction was not a key factor affecting learning performance since the computer-mouse group also improved in the post-test and delayed post-test. However, interactive games were a key factor for long-term retention, as shown by the delayed post-test ($B_3 > B_1, p < .05$), since both the computer-mouse group and motion-sensing group played the same interactive game with a questioning strategy. In addition, that the post-test results are better than the delayed post-test results in both the motion-sensing and computer-mouse group ($B_2 > B_3, p < .05$). This reveals that some loss of retention always occurs over time.

**Table 7. Simple main effect of three tests (B) with types of interaction (A)**

<table>
<thead>
<tr>
<th>Types of interaction</th>
<th>Three tests</th>
<th>$N$</th>
<th>$M$</th>
<th>$SD$</th>
<th>LSD method</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 Control</td>
<td>B1 Pre-test</td>
<td>40</td>
<td>57.20</td>
<td>3.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B2 Post-test</td>
<td>40</td>
<td>60.70</td>
<td>3.51</td>
<td>$B_2 &gt; B_1^*$</td>
</tr>
<tr>
<td></td>
<td>B3 Delayed post-test</td>
<td>40</td>
<td>59.70</td>
<td>3.45</td>
<td></td>
</tr>
<tr>
<td>A2 Computer-mouse</td>
<td>B1 Pre-test</td>
<td>40</td>
<td>54.70</td>
<td>3.58</td>
<td>$B_2 &gt; B_1^*$</td>
</tr>
<tr>
<td></td>
<td>B2 Post-test</td>
<td>40</td>
<td>67.20</td>
<td>3.56</td>
<td>$B_3 &gt; B_1^*$</td>
</tr>
<tr>
<td></td>
<td>B3 Delayed post-test</td>
<td>40</td>
<td>63.70</td>
<td>3.47</td>
<td></td>
</tr>
<tr>
<td>A3 Motion-sensing</td>
<td>B1 Pre-test</td>
<td>40</td>
<td>54.10</td>
<td>3.45</td>
<td>$B_2 &gt; B_1^*$</td>
</tr>
<tr>
<td></td>
<td>B2 Post-test</td>
<td>40</td>
<td>71.30</td>
<td>3.54</td>
<td>$B_3 &gt; B_1^*$</td>
</tr>
<tr>
<td></td>
<td>B3 Delayed post-test</td>
<td>40</td>
<td>65.60</td>
<td>3.51</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $^{*} p < .05$.

**Discussion**

In agreement with past research, the interactive game with a questioning strategy used by the two groups benefited students' long-term retention

Overall, we found a partially disordinal interaction relationship between the three groups and three tests. The post-hoc comparison found the three tests to have an order relationship (the post-tests of the three groups were better than the pre-test ($B_2 > B_1, p < .05$), showing that the three groups all had short-term learning retention; however, the control group displayed a more apparent lack of long-term retention, as the delayed post-test was not significantly superior to the pre-test. In addition, both the computer-mouse group and motion-sensing group using interactive games not only displayed significantly better performance on the post-test than on the pre-test ($B_2 > B_1, p < .05$), but also displayed significantly better performance on the delayed post-test than on the pre-test ($B_3 > B_1, p < .05$), showing that the game-based learning with a questioning strategy used for the two interactive groups resulted in improved long-term learning retention of English vocabulary. Past studies (Dale, 1969; Papastergiou, 2009; Prensky, 2005) indicated that interactive learning should result in higher learning retention. They also indicated that a questioning strategy can be applied to English learning for better retention (Basturkmen, 2001; Boyd & Rubin, 2006; Cecil & Pfeifer, 2011; Shomoossi, 2004; Yang, 2010). Research has also showed that game-based English learning promotes retention (Flores, 2015). Therefore, confirming past research, our experimental design (the football game with a questioning strategy) also promoted long-term learning retention. Since the forty English Vocabulary Words were “randomly” sampled for the pilot test regardless of how familiar they were to the subjects, this could be a reason that the passing rate of the three tests were lower than those past tests held in case classes.

**The type of interactive “interfaces” was not a key-factor affecting learning retention**

The results in this study also show that there was no significant difference between the two interactive types (three group comparison). That is, the KMIS motion-sensing interface did not outperform the computer-mouse
interface. Therefore it seems that the type of interactive “interfaces” were not the key-factors affecting learning retention. Nevertheless, previous research (Pan, 2013; Sommool et al., 2013; Vrellis et al., 2014; Yuan et al., 2015) indicated that the novelty of the Kinect motion-sensing interface can better increase students’ attention or motivation to participate when compared to a traditional computer-mouse interface did. This reveals that the novel KMIS interface perhaps could be applied to attract students’ attention or enhance motivation to participate in English vocabulary learning when compared to one-way rote memorization of English vocabulary commonly practiced in schools (Smith et al., 2013).

Since the development of applications of the Kinect in education is only just starting to expand, learners can expect much more innovation in interactive technologies. As learners adapt to motion-sensing technology, a new generation of interfaces could easily emerge for learners, similar to users’ adaptation to computer-mouse operations in the 1980s (Karat, McDonald, & Anderson, 1986). The research after two decades (Forlines, Wigdor, Shen, & Balakrishnan, 2007) has discovered that, even compared to direct touch, users were still used to using mice to operate tasks on personal computers. Motion-sensing is still relatively novel compared to learners’ familiarity with mouse operation. For the participants in our study, the computer-mouse group had the benefit of familiarity (familiarity results in lower cognitive load), but the operation was comparatively less novel (possibly leaving a more shallow impression in the memory) (Pan et al., 2014). As a result, we should consider that both the computer-mouse and the KMIS motion-sensing interface have their advantages in English vocabulary learning. The novel KMIS “interface” is helpful to attract students’ attention and motivation; however, in this study this interface was not particularly beneficial to enhancing students’ retention in English vocabulary learning.

The Kinect motion-sensing operating commands need to be intuitive in the future

In this study, participants were unfamiliar with the KMIS interface. The meanings of certain motion-sensing postures which correspond to different computer commands could differ in various research studies. This could result in unfamiliarity with operations if the definitions of postures vary across different learning environments. This also could be a hindrance in educational applications. Therefore, in further research, it would be beneficial to transform the posture definitions to make them intuitive and user friendly and avoid too complex postures to correspond to computer commands. On the other hand, further work could also consider the extent to which the KMIS approach with game-based learning can be used to learn other parts of English such as English grammar, or even other languages. This above issues are topics for future research.

Conclusion

After discussing the statistical analyses and their inferred meanings, the following conclusions were made: (1) the analyses of two-way mixed-design ANOVA reveal partially disordinal interaction relationship between the interactive types and the three tests ($F = 14.98, p < .01, \eta^2 = .204$). The post-hoc comparison (LSD method) shows that the three tests had an order relationship ($B_2 > B_1, p < .05$). This reveals that all three groups have short-term learning effects. (2) The within-group comparison, both for the motion-sensing and computer-mouse groups utilized an interactive game with a questioning strategy, displaying significant long-term (1-month) retention ($B_3 > B_1, p < .05$). Although the control group displayed a vocabulary learning effect in the post-test ($B_2 > B_1, p < .05$), there was a more apparent lack of long-term retention ($B_1$ not higher than $B_1$). (3) The between-group comparison of the two interactive groups did not reach a significant difference in English vocabulary learning, meaning that the motion-sensing interface of the KMIS was not a key-factor affecting short-term or long-term learning retention. The key-factor was the interactive content applied by the two groups.

Based on the experimental findings, our suggestion is that teachers can adopt interactive games with a questioning strategy to enhance students’ long-term English vocabulary retention. Teachers also can use the novel KMIS interface for interactive operation in order to attract students’ attention in English vocabulary learning. Learners are still relatively unfamiliar with using the Kinect interface for educational applications. This could be a hindrance in educational applications using Kinect. It would be beneficial to transform the posture definitions to make them more intuitive and user friendly and to avoid complex postures. In this study, the quasi-experimental design and cluster sampling were adopted for convenience; however, it could result in sampling error affecting experimental validity. Therefore, in future study, a true-experimental design should be adopted for better control of the interference factors.
Acknowledgements

The deepest gratitude goes to the Ministry of Science and Technology of Taiwan for its funding (MOST 102-2511-S-259-014), and the Review Panel for its utmost advice.

References


Using Novel Word Context Measures to Predict Human Ratings of Lexical Proficiency

Cynthia M. Berger1*, Scott A. Crossley1 and Kristopher Kyle2

1Georgia State University, Atlanta, GA, USA // 2University of Hawai‘i at Manoa, Honolulu, HI, USA // cberger@gsu.edu // scrossley@gsu.edu // kkyle@hawaii.edu

*Corresponding author

ABSTRACT

This study introduces a model of lexical proficiency based on novel computational indices related to word context. The indices come from an updated version of the Tool for the Automatic Analysis of Lexical Sophistication (TAALES) and include associative, lexical, and semantic measures of word context. Human ratings of holistic lexical proficiency were obtained for a spoken corpus of 240 transcribed texts produced by second language (L2) adult English learners and native English speakers (NESs). Correlations between lexical proficiency scores from trained human raters and contextual indices were examined and a regression analysis was conducted to investigate the potential for contextual indices to predict proficiency scores. Four indices accounted for approximately 42% of the variance in lexical proficiency scores in the transcribed speech samples. These indices were related to associative, lexical, and semantic operationalizations of word context. The findings demonstrate that computational measures of word context can predict human ratings of lexical proficiency and suggest that lexical, semantic, and associative context each play an important role in the development of lexical proficiency.

Keywords

Second language acquisition, Lexical proficiency, Word context, Natural language processing, Vocabulary

Introduction

While most researchers agree that lexical proficiency is an important component of second language (L2) language competence and academic achievement (Alderson, 2005; Daller, Van Hout, & Treflers-Daller, 2003; Laufer, 1992), the construct of lexical proficiency itself is still poorly understood and the field of L2 research lacks a unified theory of vocabulary acquisition (David, 2008; Schmitt, 2010). This is troubling given the need to understand how L2 lexicons develop in order to allow principled decisions regarding language pedagogy, student placement, and curricula.

Past studies investigating L2 lexical proficiency have examined the intrinsic difficulty of lexical items (e.g., Laufer, 1997), the development of lexical automaticity (e.g., Hulstijn, Van Gelderen, & Schoonen, 2009), receptive vs. productive lexical knowledge (e.g., Melka, 1997), and the distinction between breadth and depth of knowledge (e.g., Read, 2000). Another approach to conceptualizing and assessing L2 lexical proficiency examines the manner in which L2 lexical items are stored, processed, and retrieved from the mental lexicon (Aitchison, 1994). The assumption behind such an approach is that newly acquired lexical items will need to assimilate into a network of already known words, resulting in restructuring of the network as whole. Lexical proficiency is thus understood as the ability to not only differentiate between semantically related words, but to recognize the variety of ways in which lexical items may be connected to one another (Read, 2004; Singleton, 1999). Presumably, the L2 lexicon strengthens as learners develop stronger links between items and are able to more easily accommodate new words in the network (Haastrop & Henriksen, 2000). The traditional approach to investigating mental lexicons analyzes online language processing in order to gain insights into the mental lexicon and how lexical items are stored, processed, and retrieved (e.g., Conklin & Schmitt, 2008; Ellis & Beaton, 1993; Laufer, 1997). While word frequency is generally considered one of the best predictors of language processing (Balota & Chumbley, 1984; Whaley, 1978), in the current study we investigate the variability of word context to better understand lexical proficiency from a network perspective.

A promising method for better understanding L2 lexical proficiency lies in the use of natural language processing (NLP, Meurers, 2013) tools, such as the Tool for the Automatic Analysis of Lexical Sophistication (Kyle & Crossley, 2015), Coh-Metrix (Graesser, McNamara, Louworse, & Cai, 2004) and AntWordProfiler (Anthony, 2014). Such NLP analytics allow researchers and educators to operationalize many of the constructs related to lexical proficiency and quantify them in learner-produced data. The aim of the current study is to determine if computational indices related to associative, lexical, and semantic word context can increase our understanding of L2 lexical knowledge. To do so, we analyze the spoken lexical output of both L2 English learners and native English speakers in relation to human ratings of lexical proficiency using five novel indices made available in a recently updated version of TAALES (version 2.0).
The key questions that motivate our study are as follows:

- What is the relationship between word context and human ratings of lexical proficiency?
- Can measures of lexical, semantic, and associative word context predict human ratings of lexical proficiency?

Our goal is to predict human ratings of holistic lexical proficiency in spoken language to better understand the construct of lexical proficiency and the role that word context may play in identifying and predicting lexical proficiency in L2 populations.

**Background**

**Natural language processing analytics and lexical proficiency**

An emerging approach to better understanding lexical proficiency involves the computational analysis of language produced by language learners (Meurers, 2013). In such an approach, NLP tools are used to analyze large samples of learner-produced text (spoken or written) with the goal of gaining further insight into learner language and the development of lexical proficiency. Linguistic features investigated in learner output typically include psycholinguistic word properties, such as frequency and familiarity (Balota et al., 2004), age of acquisition (Ellis & Morrison, 1998), word associations (Nelson & Friedman, 1980), and imageability, concreteness, and meaningfulness (Altarriba, Bauer, & Benvenuto, 1999; Paivio, 1991). NLP approaches have demonstrated strong relationships between computational indices and human ratings of lexical proficiency or at predicting the development of lexical proficiency over time. For example, NLP studies of L2 lexis have shown that more lexically proficient L2 learners produce less concrete words, less specific words, a greater variety of words, less frequent words, less familiar words, and words with more senses (Crossley & McNamara, 2013; Crossley, Salsbury, & McNamara, 2009; Crossley, Salsbury, & McNamara, 2013; Crossley, Salsbury, McNamara, & Jarvis, 2010; Crossley, Salsbury, McNamara, & Jarvis, 2011; Kyle & Crossley, 2015).

Still, the use of computational indices to predict lexical proficiency remains incomplete, in part because NLP approaches are often limited to the analysis of individual words produced by learners (e.g., word frequency or concreteness). Such approaches do not give researchers insight into how lexical items are understood in relation to other items in the mental lexicon. One potential solution is to develop computational indices that quantify the relationship of individual words to other words in the lexicon. This approach is taken up by the current study and described further below.

**The role of context**

Historically, one of the most consistently successful predictors of lexical proficiency and growth over time has been word frequency (Balota & Chumbley, 1984; Whaley, 1978). As learners gain lexical proficiency, they use words that are less frequent in everyday usage (Ellis, 2002a; Ellis, 2002b; Sorrell, 2013). The frequency data upon which such research is based are typically word occurrences derived from large reference corpora. However, this manner of operationalizing frequency has been criticized for treating language as a randomly ordered collection of words, an approach that assumes complete independence of items in the lexicon (McDonald & Shillcock, 2001). Some researchers (e.g., Adelman et al., 2006; Brysbaert & New, 2009a) have suggested that the diversity of contexts in which a word is encountered and the subsequent constraints that context puts on a word’s meaning and use may offer more psychologically valid explanations of the word frequency effect.

Anderson’s (1991) principle of likely need presumes that the goal of human memory is to be efficient while fulfilling goals within a specific environment. Thus, items stored in memory are not equally necessary in any given context. Rather, each item has a need-probability based on past use and current context, with those items that are most likely to be needed more easily available. From this argument, if we assume that the environment in which a word is encountered is a context, we can infer that words more likely to be needed in a variety of contexts will be more readily available. This context-oriented perspective on frequency-based learning predicts that the context(s) in which an item is encountered will influence its subsequent retrieval and production. However, as previously discussed, a word’s context is not captured by the frequency measures commonly used in investigations into language acquisition. It may be that a more context-oriented construct, implicitly predicted by Anderson’s the principle of likely need, has explanatory power in this arena as well.
While the current study seeks to explore this possibility, we are not the first to do so. One approach to acknowledging the role of context in word representation has been to define context broadly as the number of documents or genres in which a lexical item is typically encountered. For example, Adelman et al. (2006) used the term contextual diversity to describe the variability of contexts (i.e., documents) in which a word occurred across three different corpora. The authors demonstrated that contextual diversity predicted word processing times in NES online psycholinguistic tasks independent of frequency and regardless of variables related to concreteness, imageability, and ambiguity. Adelman et al.’s findings were substantiated by Brysbaert and New (2009b), who offered a similar measure derived from a corpus of film and television subtitles. Brysbaert and New demonstrated that their measure (i.e., the number of films or television shows in which a word occurred) was more predictive of NES lexical decision response times than a word frequency measure alone.

The majority of work on contextual diversity has explored how L1 subjects retrieve and process words, rather than investigating L2 acquisition or lexical proficiency. In an exception, Kyle and Crossley (2015) demonstrated that approximately 26% of the variance in spoken lexical proficiency ratings could be explained by a contextual diversity index derived from the written BNC. The correlation between contextual diversity and lexical proficiency scores was negative in their study, suggesting that raters’ impressions were positively impacted by speakers who used words occurring in fewer contexts. Another approach to word context and language proficiency was taken by Crossley, Subtirelu, and Salsbury (2013), who used word association norms to investigate about 100 nouns and verbs most frequently produced by beginning-level English learners. While the authors did not find that this operationalization of context predicted the words learners produced, their basic approach is one that we adopt and expand in the current study.

Current study

In the majority of contextual diversity studies described above, context is defined as either an individual document or an entire genre. In this regard, a large-grained approach is used to conceptualize context, one that ignores a word’s lexical and semantic environments, as well as its affiliates in the mental lexicon. In the current study, we adopt the term contextual distinctiveness (McDonald & Shillcock, 2001) to refer to the constraints put on a word by its immediate lexical and semantic context. We also employ the term to address context in cognitive representation. While contextual distinctiveness was originally coined to refer to a word’s proximate lexical environment, its use in this paper is intended to more broadly reference not only lexical context, but associative and semantic word contexts as well. We do this by examining the degree to which computational indices quantifying the role of word context in language usage and cognitive representation may predict human ratings of lexical proficiency. We predict that these measures may serve to quantify the unique role that contextual distinctiveness plays in defining L2 lexical acquisition and proficiency.

We analyzed a corpus of transcribed speech samples using indices made available in the text analysis tool TAALES 2.0 (Kyle & Crossley, 2015). In order to capture a wide variety of lexical proficiency, we analyzed speech samples from 180 L2 learners at three different levels of proficiency and an additional 60 speech samples produced by native English speakers (NESs). Trained raters scored the speech samples using a holistic lexical proficiency rubric. Prior to statistical analysis, we divided the scored speech samples into a training and a test set. We then conducted correlational and linear regression analysis on the training set to examine the relations between the human lexical proficiency scores and the TAALES indices. The same model was then extended to the test set in order to cross-validate the model and ensure generalizability.

Method

Corpus

The corpus analyzed in this study contained transcribed, spoken data collected from both L2 and NESs (Crossley et al., 2010). In the L2 sub-corpus, transcribed data were derived from naturalistic, interactional discourse between an English learner and a NES interlocutor. L2 learners were either matriculated undergraduates or intensive English program students at two different universities in the United States. First language (L1) backgrounds of the learners included Arabic, French, Turkish, Japanese, Korean, Mandarin, and Spanish. Prior to participating in the study, learners were grouped into beginning (n = 60), intermediate (n = 60), and advanced (n = 60) proficiency levels based on the Test of English as a Foreign Language (TOEFL) or ACT ESL Compass scores, for a total of 180 L2 speech samples collected across levels. Proficiency levels were used to ensure a range of proficiency variance in the speech samples that comprise the corpus; however, the levels themselves
were not used as dependent variables in the current study. The NES corpus was comprised of 60 speech samples selected from the *Switchboard* corpus (Godfrey & Holliman, 1993; see Crossley et al., 2010 for details), a collection of approximately 2,400 telephone conversations taken from naturalistic conversations between 543 NESs across the United States. In total, the corpus analyzed in the current study included 240 speech samples.

**Human ratings of lexical proficiency**

Human ratings were based on transcribed interactions between a given speaker and his or her interlocutor. The corpus was divided into segments of the speech that contained approximately 150 words each for the speaker of interest and captured a complete interaction. Three trained raters assessed the 240 speech samples for lexical proficiency using a holistic grading rubric based on a 5-point Likert-scale, with a score of 5 demonstrating skillful, consistent mastery of the English lexicon and a score of 1 indicating little lexical mastery. The rubric was developed based on an adaptation of holistic proficiency rubrics produced by American College Testing (ACT), the College Board, and the American Council on the Teaching of Foreign Languages’ (Breiner-Sanders, Lowe, Miles, & Swender, 2000). In order to evaluate inter-rater reliability, Pearson correlations between all possible pairs of raters’ responses were averaged and weighted. For the full corpus, the average correlation among raters was \( r = .808 \) \((p < .001)\) with a weighted correlation of \( r = .927 \). See Crossley et al. (2010) for details regarding rubric development and rater calibration.

**Word context measures**

Five measures that take into account the role of context in word representation were included in the current study as potential predictors of L2 lexical proficiency. These are listed in Table 1 and explained further below. All the indices used were obtained using TAALES 2.0 (for details, see Kyle & Crossley, 2015). The five indices were selected to operationalize the construct of contextual distinctiveness according to associative, lexical, and semantic principles.

<table>
<thead>
<tr>
<th>Index</th>
<th>Abbreviation</th>
<th>Description</th>
<th>Subconstruct</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edinburgh Associative Thesaurus response type and token counts</td>
<td>EAT_types</td>
<td>Number of word types elicited by target word</td>
<td>Associative context</td>
<td>Kiss et al. (1973)</td>
</tr>
<tr>
<td></td>
<td>EAT_tokens</td>
<td>Number of work tokens elicited by target word</td>
<td>Associative context</td>
<td>Nelson et al., (1998)</td>
</tr>
<tr>
<td>University of South Florida stimuli counts</td>
<td>USF</td>
<td>Number of stimuli words that resulted in production of word</td>
<td>Associative context</td>
<td>Hoffman et al. (2013)</td>
</tr>
<tr>
<td>Semantic ambiguity</td>
<td>SemD</td>
<td>Variability of contexts in which word occurs</td>
<td>Semantic context</td>
<td></td>
</tr>
<tr>
<td>Relative entropy</td>
<td>McD</td>
<td>Amount of information conveyed by word about its frequent lexical contexts</td>
<td>Lexical context</td>
<td>McDonald &amp; Shillcock (2001)</td>
</tr>
</tbody>
</table>

**Associative context: Word association indices**

One approach to quantifying the distinctiveness of a word’s context is to calculate the number of other words commonly associated with it. The motivation behind incorporating word association (WA) measures in the current study was the assumption that words with a greater number of associations would be less contextually distinct. TAALES 2.0 indices derived from two publicly available behavioral datasets were used to operationalize associative context. These are described below.

**Edinburgh Association Thesaurus response type and token counts**

The Edinburgh Associative Thesaurus (EAT) index calculates the number of words produced in response to target stimuli in a written WA task. Norms are derived from NES subjects’ associative responses to 8400 English words (Kiss, Armstrong, Milroy, & Piper, 1973). Because the EAT index measures the number of associates a
word elicits, a higher number indicates a wider variety of associates. In the current study, it was assumed that words with a greater number of response associations are less contextually distinct. For example, the word worry, which elicited 65 different associate types in EAT data, may be less contextually distinct than the word husband, which elicited only 15. TAAALES 2.0 reports both type and token counts for EAT responses. Type counts reflect the number of unique words elicited in response to a given stimulus, while token counts reflect the number of responses elicited in total.

University of South Florida stimuli counts

The University of South Florida (USF) association norms (Nelson, McEvoy, & Schreiber, 1998) calculate WAs in a reverse manner from the EAT index by reporting the number of stimuli words that resulted in production of the target word as an associate in a WA task. Association data was collected from NES subjects in response to 5,019 stimulus words, resulting in a total of 10,470 response words. Words that rank high on this measure are considered more accessible and are thus more likely to come to mind in response to a variety of cues (see Nelson et al., 1998 for details). For example, the word love was produced in response to 181 different stimuli and may be relatively less contextually distinct than a word like bride, which was produced in response to only 6 stimuli in the USF WA task.

Lexical context: Relative entropy

The TAALES 2.0 relative entropy index (McD) used in the current study is derived from values calculated by McDonald and Shillcock (2001) for 8,000 English lexemes in the spoken BNC (2007). McD is based on the probability of a word co-occurring with other words in general language usage. The McD index reports the amount of information conveyed by a word about its frequent lexical contexts. A word with a higher McD value, such as lone (3.748), occurs in more distinct lexical contexts, while a word with a lower value, such as today (0.16), occurs in a variety of lexical contexts (i.e., is likely to co-occur with a variety of other words). In this example, the word lone is considered more informative about its contexts of lexical occurrence than today. In the current study, we assume that words with higher McD values are also more lexically contextually distinct.

Semantic context: Semantic ambiguity

The TAALES 2.0 semantic diversity (SemD) index operationalizes a word’s semantic ambiguity based on the variability of semantic contexts in which it occurs. This computational measure is based on Latent Semantic Analysis (Landauer et al., 1998; LSA) and was originally calculated by Hoffman, Ralph, and Rogers (2013) based on analysis of 1,000-word “contexts” in the written BNC (2007). The SemD index includes values for 31,739 English words. A high SemD value for a word is assumed to be more contextually variable, or ambiguous (i.e., occurring in a variety of semantic contexts), than a lower SemD value. For example, the word time has a SemD value of 2.30 and is thus more semantically ambiguous than the word puppy, which only has a SemD value of 0.93. The assumption motivating inclusion of SemD in the current study is that words with lower SemD values are more semantically contextually distinct than words with higher SemD values.

Analysis

Prior to conducting a stepwise linear regression, indices were checked for normal distribution. Correlations were then conducted to examine the relations between the proposed context indices and human ratings of lexical proficiency. Only indices that demonstrated a correlation of at least $r > .100$ were retained. If two or more indices demonstrated strong multicollinearity ($r > .900$), only the index that correlated the strongest with lexical proficiency would be retained. In order to cross-validate our model and assess its generalizability, we divided spoken texts into training and test sets. The training set comprised approximately 67% ($n = 166$) of the texts while the test set contained approximately 33% ($n = 74$) of the texts (Witten, Frank, & Hall, 2011). If a model derived from a training set predicts the dependent variable in a test set at a similar accuracy rate, the model can be considered stable and generalizable to the population.
Results

Correlations and normality checks

None of the indices deviated from normal distribution. All five of the indices demonstrated a correlation of at least \( r > .100 \) with lexical proficiency ratings (Table 2). No indices demonstrated multicollinearity with other indices. Thus, all five the indices were retained for regression analysis.

Regression analysis

Training set

A stepwise linear regression using the five indices yielded a significant model, \( F(4,161) = 28.713, p < .001, r = .645, r^2 = .416 \). Four indices were significant predictors in the regression: USF, SemD, McD, and EAT_types. Two of these indices had been selected to operationalize associative context (USF and EAT_types), while SemD captures semantic context and McD indexes lexical context. EAT_tokens was not a significant predictor and was not included in the model.

| Table 2. Correlations between lexical proficiency scores and word context indices |
|-----------------------------|-----|-----|
| Index          | \( r \) | \( p \) |
| USF            | -.534 | 0   |
| SemD           | .475  | 0   |
| McD            | .318  | 0   |
| EAT_tokens     | .191  | .003|
| EAT_types      | .116  | .073|

The results of the regression model (Table 3) demonstrate that the combination of four contextual indices accounts for roughly 42% of the variance in human judgments of lexical proficiency for the 166 essays that comprised the training set. The standardized coefficients (\( \beta \)) in Table 3 indicate the number of standard deviation (SD) changes we would expect in lexical proficiency for a one SD change in any given index. For example, for every one SD increase in USF value for a given speech sample, we can expect to see a 0.334 SD decrease in lexical proficiency ratings.

| Table 3. Linear regression results for contextual indices |
|-------------|-----------|-----------|-----------|-----------|-----------|-----------|
| Entry | Index | \( r \) | \( R^2 \) | \( R^2 \) change | \( \beta \) | \( SE \) | \( B \) | \( T \) |
| 1    | USF   | .502    | .252    | .252       | -.334 | 0.009 | -0.046 | -4.944 |
| 2    | SemD  | .591    | .349    | .097       | 0.350 | 1.346 | 7.207  | 5.356  |
| 3    | McD   | .630    | .397    | .048       | 0.216 | 0.550 | 1.885  | 3.430  |
| 4    | EAT_types | .645 | .416    | .019       | 0.140 | 0.032 | 0.073  | 2.297  |

Note. \( \beta \) = standardized; \( B \) = unstandardized \( \beta \); Estimated constant term is -14.732; all \( t \) significant at < .05.

Test set

The model for the test set yielded \( r = .702, r^2 = .493 \). These results indicate that the four indices retained in the training set accounted for roughly 49% of the variance in human ratings of lexical proficiency for the 74 texts that comprised the test set.

Discussion

The purpose of this study was to investigate links between novel computational word context indices and human ratings of lexical proficiency. The results indicate that there are strong relationships between contextual distinctiveness and lexical proficiency. The findings have important implications for second language acquisition (SLA) and lexical proficiency because they suggest that lexical, semantic, and associative context may each contribute to the development of lexical proficiency.

Our first research question asked whether there was a relationship between word context and human ratings of lexical proficiency. The findings suggest that there is indeed such a relationship, with all contextual
distinctiveness indices showing significant correlations with lexical proficiency (Table 2). USF demonstrated a large effect size ($r > .50$) in its correlation with lexical proficiency. In addition, two other indices (SemD and McD) demonstrated medium effect sizes ($r > .30$), while two (EAT_tokens and EAT_types) demonstrated small effect sizes ($r > .10$) (Cohen, 1988).

Interestingly, an inverse relationship between semantic ambiguity (SemD) and lexical variability (McD) was reported: As speakers become more lexically proficient, they use words that are more semantically ambiguous but also more lexically distinct. Thus, more proficient speakers use words that occur in a wide variety of semantic contexts (higher SemD) but are less likely to co-occur with a wide variety of words (higher McD). This finding suggests that semantic and lexical context are distinct subconstructs that may capture unique aspects of contextual distinctiveness. There was also an inverse relationship between EAT response type and token counts and USF stimuli counts. Recall that both sets of WA measures were selected to operationalize associative word context (i.e., how words are related to one another in the mental lexicon): The former reports the number of response to a target word while the latter reports stimuli counts. Because the relationship between USF and lexical proficiency was significantly stronger than the relationship between EAT and lexical proficiency (Tables 2 and 3), it could be that stimuli counts (e.g., USF norms) represent a more robust manner of quantifying associative context than the analysis of response counts. The existence of inverse relationships among indices suggests that each measures distinct subconstructs of contextual distinctiveness. This explanation is elaborated below when we analyze speech samples with reference to the four indices.

Our second research question asked whether contextual distinctiveness measures could predict human ratings of holistic lexical proficiency. Indeed, a combination of four indices related to word context (Table 3) explained nearly 42% of the variance in lexical proficiency in transcribed speech samples. This model retained indices pertaining to each operationalization of word context: associative, lexical, and semantic. Relationships between lexical proficiency and the indices selected by our model are outlined in Table 4 and further explained below.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Subconstruct</th>
<th>Variance Explained</th>
<th>More Proficient Speakers…</th>
</tr>
</thead>
<tbody>
<tr>
<td>USF</td>
<td>Associative context</td>
<td>25%</td>
<td>Used words associated with a smaller number of stimuli in WA tasks</td>
</tr>
<tr>
<td>SemD</td>
<td>Semantic context</td>
<td>10%</td>
<td>Used words that occur in a wider variety of semantic contexts</td>
</tr>
<tr>
<td>McD</td>
<td>Lexical context</td>
<td>5%</td>
<td>Used words that occur in more distinct lexical contexts</td>
</tr>
<tr>
<td>EAT_types</td>
<td>Associative context</td>
<td>2%</td>
<td>Used words that produce a variety of responses in WA tasks</td>
</tr>
</tbody>
</table>

The greatest predictor of lexical proficiency in speech samples was USF stimuli counts, which accounted for over 25% of the variance in human ratings. The correlation between USF and lexical proficiency ratings was negative, suggesting that as speakers gain lexical proficiency they use words that are less likely to be produced in response to a range of stimuli in free association tasks. The second strongest predictor of lexical proficiency was SemD, which explained another 10% of variance in ratings. SemD was positively correlated with lexical proficiency, meaning that more lexically proficient speakers produce words that occur in a wider variety of semantic contexts. The third strongest predictor of lexical proficiency, McD, explained roughly 5% of lexical proficiency scores. McD reports the amount of information conveyed about a word’s frequent lexical contexts in language usage, with a higher value indexing words that statistically co-occur in the company of fewer words. Because this measure positively correlated with ratings, we know that more proficient speakers are more likely to produce words that occur in more distinct lexical contexts. The final predictor in our model, EAT_types, explained 2% of the variance in holistic lexical proficiency ratings. This index was positively correlated with lexical proficiency and suggests that more proficient speakers produce words that elicit a greater variety of responses in word association (WA) tasks.

To demonstrate how the above model reflects lexical proficiency in actual language use, we offer a comparison of two contrasting speech samples from the corpus. Sample A received the highest rating possible from human raters, while Sample B received one of the lowest. Certain words have been italicized to highlight their reference in discussion below.
Sample A
Okay. I don’t really, I more, I don’t know about the government as much as the people. I wouldn’t consider to be a threat at all and I really don’t feel much like the Soviet Union itself is a threat anymore. I’m, I’m worried about them. They’re in a very tumultuous state right now with the kinds of adaptations that they’re attempting to go through. Yeah. Yeah I think that’s… that’s a real important aspect and that as the… the… the most the mo… let’s see, the more that we do that we do or that we can do to help them become self-sufficient is going to eliminate more of the risk of that becoming, you know, a reality. I know that this last winter was very hard on several areas in the… in the… the Ukraine, particularly the coal mining regions of Siberia. The people there have money that’s not their problem, but there’s no food for them to buy and it’s, you can’t eat money.

Sample B

Table 5 reports the speech samples’ holistic proficiency ratings, as well as their scores for each of the four indices retained by our model. Index values are reported in z-scores, which adjust a given distribution to have a mean of 0 and a standard deviation of 1. Figure 1 represents these z-scores visually.

<table>
<thead>
<tr>
<th>Speech sample</th>
<th>Lexical proficiency rating</th>
<th>USF</th>
<th>Sem_D</th>
<th>McD_CD</th>
<th>EAT_types</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5.000</td>
<td>-0.545</td>
<td>1.596</td>
<td>1.368</td>
<td>1.438</td>
</tr>
<tr>
<td>B</td>
<td>1.667</td>
<td>0.715</td>
<td>-2.115</td>
<td>0.242</td>
<td>-0.329</td>
</tr>
</tbody>
</table>

Figure 1. Z-scores for speech Samples A and B

The two speech samples analyzed here performed as predicted by our model, with the exception of McD_CD in Sample B, which was slightly above the group mean (rather than below, as predicted). However, it was still lower than Sample A’s McD_CD score. Because USF correlated negatively with lexical proficiency in our model, the highly proficient speaker that produced Sample A used fewer words elicited by a range of stimulus words in the USF WA task. For instance, Sample A contains words such as aspect (1 USF) and reality (6 USF), while sample B used more words elicited by a range of different stimulus words, such as school (183 USF) and class (95 USF). Unlike USF, EAT types demonstrated a positive relationship with lexical proficiency. For this reason, the more highly rated Sample A contained more words that elicited a greater number of response types in the EAT WA task. This finding runs counter to the negative relationship between USF and lexical proficiency. Most likely, this is because USF and EAT measure different aspects of lexical association: EAT reports the number of responses to a target word while USF reports the number of stimuli that elicited the word. For example, reality, used in Sample A, has a low USF value (6) but a moderately high EAT_types value (43). In
other words, reality is unlikely to be elicited by a variety of stimuli in WA tasks, even though it elicits 43 different response types when used as a stimulus. Sample B, on the other hand, contains words like one, which has a high USF value (63) but a relatively low EA T_type value (23). Meanwhile, other words used in Sample B, such as school, have moderate EA T_type values (56) but extremely high USF values (183). This is because a word like school is elicited by a number of different stimuli in WA tasks but only produces a moderate number of responses when acting as a stimulus.

Table 5 and Figure 1 also indicate that Sample A had higher SemD and McD values than Sample B, as predicted by our model, suggesting that more proficient speakers produce words that are more semantically diverse at the same time that they are more lexically distinct. For example, Sample A contains eliminate, which has both a high SemD value (2.09) and a high McD value (1.787). While eliminate is likely to occur in a variety of different semantic contexts, it is more constrained in terms of words that it is likely to co-occur with in general language usage. Sample B, however, contains words like manager, which occurs in more distinct semantic contexts (SemD = 1.68) but is less lexically distinct (McD = .604).

Table 5 and Figure 1 also indicate that Sample A had higher SemD and McD values than Sample B, as predicted by our model, suggesting that more proficient speakers produce words that are more semantically diverse at the same time that they are more lexically distinct. For example, Sample A contains eliminate, which has both a high SemD value (2.09) and a high McD value (1.787). While eliminate is likely to occur in a variety of different semantic contexts, it is more constrained in terms of words that it is likely to co-occur with in general language usage. Sample B, however, contains words like manager, which occurs in more distinct semantic contexts (SemD = 1.68) but is less lexically distinct (McD = .604).

The associative, lexical, and semantic context indices in our model each explained unique variance in the ratings of lexical proficiency. In addition, none of the indices were highly inter-correlated (the highest correlation was between USF and SemD: $r = -.337$, $p > .001$). These results demonstrate that the measures introduced in our study capture distinct aspects of word context and offer initial evidence to support their independence as subconstructs of word context. This finding also indicates that the development of lexical networks may be impacted independently by the semantic contexts of newly acquired words, the associative relatedness between known and newly acquired words, and the number of statistically frequent lexical neighbors that occur in the company of a given word in everyday language use.

Our results support Kyle and Crossley’s (2015) findings, who demonstrated that a contextual diversity measure (the number of individual documents in which a word was found in a reference corpus) predicted roughly one-quarter of the variance in holistic ratings of spoken lexical proficiency. However, the approach used in Kyle and Crossley examined context globally (across texts) and not locally (within a small window of words). Only one other study, to our knowledge, has taken this approach (Crossley, Subtirelu, & Salsbury, 2013), and our findings run contrary to theirs (i.e., USF norms were not predictive of the words produced by beginning-level English learners). This likely results from Crossley, Subtirelu, and Salsbury (2013) only analyzing the most frequently produced nouns and verbs in their learner corpus.

Our findings have the potential to offer insights into L2 testing and assessment, including the automatic scoring of language skills. They indicate that word properties beyond word frequency account for human judgments of lexical proficiency. The findings also suggest that L2 instructional approaches may benefit from methods that move beyond the study of words in isolation to include word context, with emphasis on associative relationships among words and the frequent semantic and lexical contexts of language in use. For instance, one pedagogical activity that our results support is semantic word mapping, which encourages learners to brainstorm and diagram associations between words (Johnson & Steele, 1996; Laufer, 1990). Another approach to incorporating context in language teaching would involve the use of corpus-based concordance tools, which allow learners to observe target words in their most common lexical and semantic environments (Reppen, 2010; Römer, 2008).

**Conclusion**

The results of this study suggest that computational indices quantifying word context can be used to predict human ratings of spoken lexical proficiency. Our analysis of a corpus of transcribed speech samples included five indices made available in TAALES 2.0. The indices were selected to reflect the construct of word context according to associative, lexical, and semantic operationalizations. Four of these indices demonstrated a significant relationship to holistic ratings of spoken lexical proficiency. A model obtained in stepwise linear regression explained 42% of the variance in human ratings, with indices pertaining to each operationalization of lexical proficiency contributing to the model. Results suggest that computational measures of lexical, semantic, and associative context each play an important role in understanding lexical proficiency.

While our model has predictive validity, the indices themselves cannot be interpreted as contributing to lexical proficiency per se. Lexical proficiency itself is undoubtedly explained by a number of factors. For example, variables related to rhetorical organization, pragmatic and content knowledge, and accuracy undoubtedly impact impressions of lexical proficiency, and these are not included in the current research. Follow-up studies are needed in order to determine the causal role the proposed indices may play in explaining lexical proficiency and
the degree to which these context-related indices relate to other variables known to impact lexical proficiency. Additionally, while we have made claims about the development of proficiency, the current study is based on a cross-sectional dataset. The use of these same computationally derived context measures on a longitudinal dataset would offer greater insights into the relationship between contextual distinctiveness and the development of lexical proficiency in learners over time (Ortega & Iberri-Shea, 2005).

Nonetheless, the novel computational indices proposed in the current study offer a unique manner with which to investigate the role of contextual distinctiveness in the mental lexicon (associative) and in language use (semantic and lexical). As tools for the enhanced application of learner analytics, they also may allow researchers and educators to better analyze learner-produced data in order to assess learner progress and predict future performance.

References


EssayCritic: Writing to Learn with a Knowledge-Based Design Critiquing System

Anders I. Mørch1*, Irina Engeness1, Victor C. Cheng2, William K. Cheung2 and Kelvin C. Wong2

1Department of Education, University of Oslo, Oslo, Norway // 2Department of Computer Science, Hong Kong Baptist University, Hong Kong // anders.morch@iped.uio.no // irina.engeness@iped.uio.no // victor@comp.hkbu.edu.hk // william@comp.hkbu.edu.hk // kckwong@comp.hkbu.edu.hk

*Corresponding author

ABSTRACT

This article presents a study of EssayCritic, a computer-based writing aid for English as a foreign language (EFL) that provides feedback on the content of English essays. We compared two feedback conditions: automated feedback from EssayCritic (target class) and feedback from collaborating peers (comparison class). We used a mixed methods approach to collect and analyze the data, combining interaction analysis of classroom conversations during the writing process and statistical analysis of students’ grades. The grades of students in both classes improved from pre-test to post-test but in different ways. The students in the target class included more ideas (content) in their essays, whereas the students in the comparison class put more emphasis on the organization of their ideas. We discuss our findings to identify strengths and weaknesses of our approach, and we end the paper by suggesting some directions for further research.

Keywords

Automated feedback, Collaboration, Decision tree learning, Design critiquing framework, English as a foreign language (EFL), Essay writing, Learning analytics (LA), Machine learning, Methods for using LA in EFL, Peer feedback

Introduction

Three components in successful foreign language learning according to the call of this special issue are: learners, goals, and contexts. We address them in this paper by formative assessment (process feedback). We focus on the communicative aspects of foreign language learning, writing and speaking, in terms of content (meaning or ideas) and organization of shorter texts, and not on grammar and spelling (acquiring linguistic competences). This is an under-researched theme in language learning and contemporary research explores semantic analyses tools, which we give an example of in this article. It addresses a challenge in many countries today that young people are overexposed to learning spoken English through new media without being activated by reading and writing. The overreliance on speaking for developing language skills may have a negative effect on vocabulary development (Weizmann & Snow, 2001). We address the discrepancy by a computer-based writing aid together with collaboration in small groups, which is informed by a theoretical framework for integrating action (writing) and reflection (thinking aloud or talking), the design-critiquing framework.

Learning analytics (LA) is a relatively new concept in the learning sciences, but it has existed under other names for more than 30 years in smaller scale. LA refers to the analysis of the learning process, in particular the traces of learning that can be captured by tools and adaptive teaching methods. LA data are dependent on observable data collected from any learning episode, including but not limited to educational technology and learning management systems. LA researchers also investigate impacts of learning traces on administrative policy and curriculum reform. Through analytics, institutions (e.g., universities, schools, online education providers) can collect large data sets and apply statistical techniques to predict success or failure, and give advice and suggestions. This may be through informing instructors how specific students are struggling so that they can contact those learners with advice (Baker & Siemens, 2014) or by informing the students directly by presenting automated feedback in the user interface of the educational technology (Fischer et al., 1991). LA departs from purely technical approaches that use data as their sole resource for analysis (educational data mining) and often involve theory to guide the selection of research methods, design of the educational technology, and the interpretation of usage data (Baker & Siemens, 2014).

Literature review

Feedback during the writing process (formative assessment) is an essential component of the teacher’s role in English writing classes and is used to improve students’ writing skills (Hattie & Timperley, 2007; Black & Wiliam, 2009). Previously, Black and Wiliam (2009) identified the following types of formative assessment in
English: (1) clarifying and sharing learning intentions and criteria for success, (2) engineering effective classroom discussions and learning tasks that elicit evidence of student understanding, (3) providing feedback that moves learners forward, and (4) activating students as instructional resources for one another and as owners of their own learning. Moreover, the effective teacher when giving feedback should address three major issues in a student’s learning process (Hattie & Timperley, 2007): (1) Where am I going? (What are the goals?) (2) How am I doing? (What progress is being made toward meeting the goals?) (3) Where to next? (What activities need to be undertaken to make better progress?) Thus, effective feedback is able to bridge the gap between students’ prior knowledge and the new knowledge encapsulated in a learning assignment or a learning goal.

Peer assessment is an alternative type of formative assessment often used in educational practice to stimulate student collaboration on texts in progress (Birenbaum, 2003). Previous studies on EFL learning have compared teacher feedback with peer feedback, and researchers have found that teacher feedback is more likely to be incorporated in redrafts than peer feedback (e.g., Yang, Badger, & Yu, 2006). However, to the best of our knowledge, very few studies have compared peer feedback and automated (computer based) feedback.

Research on computer-based feedback includes automated essay scoring (AES). AES systems assign scores to essays written for educational purposes. The score is dynamically computed by machine learning and statistical techniques on large data sets, often based on supervised learning algorithms (Hastie, Tibshirani, & Friedman, 2001). There are disagreements in the literature about the strengths and weaknesses of automated assessment (Foltz, 2014; Kukich, 2000; Lee et al., 2009; Mørch et al., 2005; Sireci & Rizavi, 2000). Proponents have argued that these digital aids successfully compare with the accuracy and reliability of human evaluation (Sireci & Rizavi, 2000). However, critics have pointed out that such systems do not encourage students to pursue novelty in their writing and instead lead to conformity; furthermore, these systems can be fooled by intentional gibberish and thus can give students inaccurate, higher scores (Kukich, 2000).

We aimed to fill a niche in the previous research and designed an experiment to address the following research questions: How do the two types of feedback (automated vs. peer) assist EFL students during the writing process, and what impact does the feedback have on their final essay grades?

**Design critiquing framework**

The design-critiquing framework (DCF) is a theoretical framework that integrates action (doing) and reflection (conversing and thinking) within a computational environment (Fischer et al., 1991; Robbins & Redmiles, 1998; Mørch et al., 2005). It was inspired by Donald Schön’s theory of the reflective practitioner (Schön, 1983), which suggests that skilled performers toggle between action and reflection, as in thinking about alternatives and what to do next when they create. Novice designers need scaffolding to integrate action and reflection and this is provided by “back talk” from the environment (e.g., problematic situations, contingencies, nudges, hints, prompts) to trigger the shift from action to reflection (Schön, 1983). Within a computational environment the shift can be accomplished by automated feedback generated from domain-specific rules (Fischer et al., 1991). A goal of “back talk” is to bringing a degree of objectivity to the novices’ highly subjective creative process of designing. Full computational support of DCF requires automated analysis of artifacts and deliberations. The current version of EssayCritic (version 3) supports artifact analysis.

The computer’s role in DCF in this study is to analyze textual artifacts (essays) and make suggestions to students for improving the text, according to a writing assignment. Hattie and Timperley (2007) argue effective feedback connects students’ prior knowledge with the desired knowledge. With DCF, this means using automated feedback to intervene in the action-reflection loop and create a shift from action to reflection, triggering talk (reflection) based on the current version of an essay (action_n), which may lead to subsequent revision (action_{n+1}). The writing process ends when the assignment has been completed to a sufficient degree or when the allocated time is up. We explored how EFL students used feedback produced by EssayCritic in two ways: (1) in small group discussions triggered by the feedback (surfacing students’ prior knowledge) and (2) feedback incorporated as new sentences in their essays (approximating desired knowledge).

**The EssayCritic system**

The EssayCritic system is a web application for semantic analysis of short texts (< 500 words) (Lee et al., 2009; Mørch et al., 2005). The key feature of EssayCritic is its identification of the presence or absence of subthemes (specific topics/ideas written about) in individual essays. Its novelty is that it decomposes the subthemes into
simpler concepts and this makes concept identification, and subsequently topic/idea identification, easier and more accurate.

In preparing the knowledge base for a new essay topic, the first step is to create a concept tree representing the topic; this is done in collaboration between the teachers and researcher. The teachers suggest a set of desired subthemes for the topic, obtained from the analysis of the content of a textbook chapter, and provide a set of sample essays collected from previous students. The researcher decomposes each subtheme into simpler concepts, which can be precisely represented by a few phrases or keywords using synonyms from dictionaries, WordNet (Fellbaum, 1998), textbooks, and texts from the sample essays acquired from students, newspapers, and web pages.

During the system-training phase, the second step, the researcher manually labels the sentences of the sample essays to identify which concepts are contained in them. These instances act as the training data for the decision tree algorithms (Quinlan, 1986), which output a set of rules, i.e., logic combinations of the derived simple concepts, indicating the relationship between these concepts and the subthemes (Appendix A). In this study, eleven subthemes were identified for the essay assignment (Appendix B). The overall time required to prepare the knowledge base and train the system was approximately four weeks. Most of the time is spent on preparing the knowledge base. Training the system took about two to three days, including fine-tuning of the rules. In terms of supervised machine learning, the system-training phase is very short compared to many other applications.

The students invoke EssayCritic by uploading essays to a server that are examined sentence by sentence. If a sentence is found to satisfy a rule, it is presumed to contain the associated subtheme, and information can be presented as feedback in the user interface (Figure 1). The students can choose between two types of feedback: (1) covered subthemes (Figure 1: left) and (2) suggested subthemes (Figure 1: right).

**Figure 1.** The user interface of EssayCritic: Covered subthemes (left) and absent (and suggested) subthemes (right)

Research design and mixed methods

Five teachers of a Norwegian upper secondary school created the following assignment given to 125 students aged 16-17 from five classes: “Write an essay on the topic of English as a global language. Explore how English was spread around the globe, and present the most important reasons for this development. About 300-400 words.” All students wrote the first draft using a word processor, and two teachers graded the essays by using an
assessment rubric aimed at evaluating the content of students’ essays. Eleven best essays from three classes were selected to train the system. The remaining two classes of 24+24 students served as target and comparison classes. Students in the target class received feedback from EssayCritic, and the students in the comparison class provided feedback to each other. The students in both classes had Norwegian as a native language; they had been studying English for 10 years and followed similar teaching plans prior to the intervention. To make the comparison fair in terms of equal access to learning resources, all students had access to the same list of the eleven EssayCritic subthemes (Appendix B) and the assessment rubric. The students in both classes were asked to focus on content and use the resources they had available during the writing process, but it was only the target class students who received the feedback integrated with their essays in the user interface of EssayCritic (see Figure 1). The students in both classes revised their essays twice and handed version 3 for grading by an independent teacher.

The students of both classes worked in small groups of four. The purpose of these groups for the target class was to collaboratively discuss the meaning and implications of the EssayCritic feedback, and allowed researchers to observe students’ deliberations. Figure 2 illustrates the setting in the target class. In the comparison class, the students first read each other’s essays and then gave oral feedback. Using this experimental set-up the research team aimed to make a fair comparison of two forms of giving feedback on written work, and to increase our understanding of the role of feedback in bridging students’ prior and desired (teacher specified) knowledge.

To analyze the data, we applied mixed methods, combining quantitative and qualitative data sets (Creswell, 2012). A paired t-test was used to analyze the improvement in grades from pretest to posttest and the improvement in the number of subthemes from the pretest to the posttest. Cohen’s d (Cohen, 1992; Field, 2013) was calculated to evaluate the effect size of these improvements. We used an independent t-test (Cohen, Manion, & Morrison, 2011; Field, 2013) to calculate the statistical difference between results of the pre- and post-tests and between the number of subthemes in the pre- and post-tests in both classes.

We also counted the number of times the teacher intervened in the two classes. In the target class the teacher intervened six times and in the comparison class 13 times. Interventions were mainly in response to students’ requests for information about the task, the ideas they talked about in groups, and if they had understood the feedback procedures.

We used three cameras to record the activity of three (out of six) groups from each class to collect qualitative data. A fourth camera followed the teacher. In total we have nine hours of video data. Field notes taken during class observations were used to contextualize the video data (Derry et al., 2010). All students communicated in English. We analyzed the conversations that occurred in two groups, one from each class, using a qualitative approach, inspired by the interaction analysis method (Jordan & Henderson, 1995). The groups were selected because the students were verbally active. We selected only two data extracts for inclusion in the analysis below due to space limitations. The transcript notation used to reproduce the verbal protocols is shown in Appendix C.

The combination of quantitative and qualitative methods allowed us to compare achievements at different levels of detail: final essays (grades and number of subthemes) and writing process (what the students were talking about).
Results and analysis

Quantitative data

We compared the results of versions 1 and 3 of the essays from both classes. We considered version 1 as a pre-test and version 3 as a post-test. In total, 96 essays marked by the classroom and independent teacher on a scale of 1-6 constituted our quantitative data. The inter-rater reliability (Cohen’s Kappa) between the grades by the classroom teachers and the independent teacher has been computed (Table 1).

Table 1. Reliability for two coders (classroom and independent teacher) of pre-test and post-test for the target and comparison classes (calculator used: http://dfreelon.org/utils/recalfront/recal2/)

<table>
<thead>
<tr>
<th>Class</th>
<th>Cohen’s Kappa, pre-test</th>
<th>Cohen’s Kappa, post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target class</td>
<td>0.226</td>
<td>0.229</td>
</tr>
<tr>
<td>Comparison class</td>
<td>0.318</td>
<td>0.296</td>
</tr>
</tbody>
</table>

Measured on the scale 0.2-1.0 (Field, 2013) the obtained results of Cohen’s Kappa indicate fair inter-rater reliability between the assessments of the classroom teacher and the independent teacher.

The results of the pre- and post-tests (average grades) are presented in Table 2.

Table 2. Average pre-test and post-test grades in the target and comparison classes and the observed groups

<table>
<thead>
<tr>
<th>Population</th>
<th>Avg. grade pre-test (essay v.1)</th>
<th>Avg. grade post-test (essay v.3)</th>
<th>Difference</th>
<th>Paired t-test</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target class (N=24)</td>
<td>2.8</td>
<td>4.5</td>
<td>1.7</td>
<td>t(24) = -7.62</td>
<td>d = 2.26</td>
</tr>
<tr>
<td>Observed group</td>
<td>2.7</td>
<td>4.5</td>
<td>1.8</td>
<td>p &lt; .0005</td>
<td></td>
</tr>
<tr>
<td>Comparison class (N=24)</td>
<td>2.9</td>
<td>4.8</td>
<td>1.9</td>
<td>t (24) = -8.86</td>
<td>d = 2.64</td>
</tr>
<tr>
<td>Observed group</td>
<td>3.5</td>
<td>4.5</td>
<td>1.0</td>
<td>p &lt; .0005</td>
<td></td>
</tr>
<tr>
<td>Independent t-test</td>
<td>p = .725</td>
<td>p = .903</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The paired t-test (Cohen et al., 2011) shows a significant difference between the average grades of the pre- and post-tests in both classes, and the Cohen’s d (Cohen, 1992; Field, 2013) indicates a large effect in both classes, as d-values larger than 1.0 is considered large according to Plonsky and Oswald (2014). The independent t-test (Field, 2013) shows no significant statistical difference between the pre-test grades and no significant statistical difference between the post-test grades (p = .725 and p = .903).

We used another measure to identify differences in outcome between the two feedback conditions (Table 3).

Table 3. Average number of subthemes in the essays of the target and comparison classes and the observed groups

<table>
<thead>
<tr>
<th>Population</th>
<th>Avg. # subthemes in pretest (essay v.1)</th>
<th>Avg. # subthemes in posttest (essay v.3)</th>
<th>Difference</th>
<th>Paired t-test</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target class (N=24)</td>
<td>3.25</td>
<td>7</td>
<td>3.75</td>
<td>t(24) = -9.70</td>
<td>d = 4.72</td>
</tr>
<tr>
<td>Observed group</td>
<td>4</td>
<td>8.25</td>
<td>4.25</td>
<td>p &lt; .0005</td>
<td></td>
</tr>
<tr>
<td>Comparison class (N=24)</td>
<td>4.04</td>
<td>5.92</td>
<td>1.88</td>
<td>t(24) = -6.91</td>
<td>d = 2.07</td>
</tr>
<tr>
<td>Observed group</td>
<td>4.5</td>
<td>6</td>
<td>1.5</td>
<td>p &lt; .0005</td>
<td></td>
</tr>
<tr>
<td>Independent t-test</td>
<td>p &lt; .0005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3 presents the average number of subthemes in the essays of both conditions. The paired t-test indicates a significant difference between the numbers of subthemes from pretest to posttest in both classes. Cohen’s d reflects a large effect in both classes. In addition, the independent t-test shows a significant statistical difference between the increase of subthemes across the two classes (p < .0005). In other words, the students in the target class included significantly more subthemes in their final essays than the students in the comparison class.

In summary: The quantitative data showed no significant difference between the final grades in the two classes, but there was a significant difference in the number of subthemes. To analyze the results in more detail we performed a qualitative analysis of the writing process.
Qualitative data

Extract 1: Analysis of the interactions between students in the target class (automated feedback)

The extract below (Extract 1) was taken from a discussion between four students. They have uploaded the first version of their essays to EssayCritic. It is Jane’s turn to share the feedback she received with the rest of the group, which included: “Say more about: Good English helps people to find high quality jobs.”

1. Jane: OK and then I have to say about how English helps people to find high quality jobs. Well, I think high quality jobs are international, in business, and in companies. Many companies deal outside the country. My dad, he works with furniture, he has to travel to Asia and Europe and he has travelled so many times. And he knows what furniture we are going to sell in Norway, and he has to know English very well because he talks to Chinese people and Japanese people.

2. Carol: Well, I went to chiropractor, and only the secretary was Norwegian. All the doctors there spoke English; they were Australians and British and Americans.

3. Jane: So, if you are going to have a job, you have to have some knowledge…

4. Carol: Yes, basically you don’t get a good job if you don’t speak English.

5. Jane: Yes, you have to know some basic English; you have to know English words for the things that you work with.

6. Carol: Yes, at least.


((Later in the writing process))

8. Teacher: Have you got any questions on the feedback on your draft?

9. Carol: I have covered seven subthemes, and I need two more.

10. Jane: Same here. I covered three when I submitted my first draft, and now I have covered seven.

11. Teacher: Oh, that’s very well. (Jane invokes the covered subtheme function of EssayCritic)

12. Jane: But I think the structure of my text is bad.

13. Teacher: Take a look at the structure as well.

The extract shows that the brief feedback was meaningful to the learners and triggered memories (lines 1-2). The girls conclude that knowledge of English is very important for working life (lines 3-7). Later in the conversation the teacher offers assistance (line 8), and the two girls report the number of subthemes they have included (lines 9-10). The teacher acknowledges with praise that they have written about most of the subthemes, but Jane is unsure about her essay’s organization (“I think the structure of my text is bad,” line 12). Her concern is raised after invoking EssayCritic’s covered subthemes function (Figure 1: left), which shows that the text for one of the subthemes is scattered throughout her essay. The teacher suggests that Jane works more on the structure but without providing specific guidance.

Jane included the following text in version 2 of her essay, addressing the subtheme “English helps people to find high quality jobs”: “We use English everywhere, and we are surrounded with it. We use English in science and technological innovations and to get a good job.” Carol wrote, “You can hardly get a high quality job, if you don’t speak English well. Sometimes you can get a high quality job, but you have to speak at least a little English, just to know what and whom you are working with.”

In summary: The feedback from EssayCritic to the target group prompted the students to write more idea-rich essays, but the focus on content creation took time away from organizing their essays for readability.

Extract 2: Analysis of the interactions between students in the comparison class (peer feedback)

In Extract 2 another group of four students have received grades on version 1 of their essays from the teacher but without any feedback. Josh and Mike have read each other’s essays, and Josh is in the middle of giving feedback to Mike.

1. Mike: We wrote mostly entire essay about history. It’s easy to fix it; we have to just make this part bigger and that part smaller ((pointing to text on computer screen)). This is an introduction; and that is finish.

2. Josh: And there is also a problem: different reasons in one paragraph, because often you should have one reason per paragraph.
3. Mike: I hate this, I had to put two reasons and I hate it. In my next draft I will do something about that. How do you feel about language to write about the reason number three? ((Referring the list of subthemes on paper))

4. Josh: You should use sentences binders.

5. Mike: Yes, I should, but I didn’t use it. It’s like every test I have. If we are allowed to use something, I always bring a pile of helping stuff and I never use it. I don’t know why.

6. Josh: I really hate giving feedback because it’s so difficult.

7. Mike: Yes, it is, even if you are a teacher.

8. Josh: And it’s worse with us. She said to focus on content, the flow...

9. Mike: I have looked at yours, and it’s very good. There is very little to correct.

10. Josh: I will try to cut it down.

11. Mike: Try to cut it down. It was too long, but this was not a problem. Maybe focus a bit more on other things than history.

((Later in the writing process))

12. Josh: If you want, you can write about the upper class situation with learning English.

13. Mike: It isn’t coming naturally here; that’s the point. And that’s about history, and I don’t want more history.


15. Mike: Because ((takes the assessment rubric)) I have already five or six sentences about history, I don’t think I need more.

16. Josh: It’s so difficult not to talk about history because the whole book is about history ((referring to the textbook)). Have you talked about music?


18. Josh: You can specify, but that’s not needed really.

19. Mike: Then there will be producing of media, music, and films and TV series ((laughing)) I feel like I want to write “entertainment.” It feels like it’s easier.

20. Josh: Yes, that’s more descriptive.

21. Mike: OK. But I feel if I want a better grade, there is something more I have to change.

22. Josh: But you can see it?

23. Mike: And I don’t know what it is. That’s kind of annoying.

Josh indicates there are some problems with the organization of Mike’s essay (line 2). The two boys are uncertain whether they should divide a paragraph in two (line 3-4). In lines 6-12, Josh expresses his frustration about the difficulty of giving feedback (“I really hate giving feedback because it’s so difficult”) and Mike ends up saying Josh’s essay is good with little to correct (line 9). Mike consults the assessment rubric (line 15). He realizes that he has written 5-6 sentences “about history” (line 15). Josh mentions “music” (line 16), and Mike says that he has written about media and entertainment but is unsure if he also needs to mention music. By elaborating on each other’s thoughts, the two peers contribute to the development of a common understanding of the assigned topic of English as a global language, albeit in a somewhat arbitrary way. They are not at all sure if they are moving in the right direction (lines 21-23).

Mike included the following in his final draft: “... the US has taken the role of a mass producer of media and entertainment in the world. After the Second World War, Europe was in ruins. The European industry and prosperity was dramatically slowed down, while in the US, the economy grew. Making the US the first and the only international superpower. Now, new Hollywood movies are displayed on the big screen all over the world, every month.” Josh wrote in his essay, “With all new smartphones and streaming possibilities of films and music, we are hearing English more than we used to. With streaming programs like Netflix and Spotify, you can watch Titanic or listen to The Beatles on the bus. The massive information we receive from international news pages also influences us.” Mike chose to combine two subthemes in his essay, media and entertainment in a historical perspective, whereas Josh wrote about the opportunities offered by streaming English language media.

**General discussion**

We were surprised at first to find that both classes received approximately the same grades on the post-tests. We did aim for a fair distribution of learning resources (all students had access to the 11 subthemes and the assessment rubric), but we anticipated that the students who received automated feedback would improve their grades on the basis of producing more content rich essays. The qualitative differences provide a clue to why the grades did not differ more.

219
Following Black and Wiliam (2009), we can state that the feedback from EssayCritic enabled the learners in the observed target group to move forward, as the feedback prompted them to “say more about” a missing subtheme and improve their essays by including more content related to the assignment. This effect was not observed to the same extent in the comparison group as we elaborate below.

Hattie and Timperley (2007) suggested that effective feedback should address the discrepancy between prior and desired knowledge. Our findings suggest that the students in the target group were more successful at building bridges. Carol and Jane were able to connect their understanding of English as a global language with personal experiences (e.g., “My dad, he works with furniture, he has to travel to Asia and Europe...” and “I went to chiropractor, and only the secretary was Norwegian, all the doctors there spoke English.” lines 1-2 in Extract 1).

In their attempts to connect their ideas with the assignment, Josh and Mike were only able to think of what the teacher would say to them when giving feedback, i.e., referring to what was expected of them. It seems that the type of learning that occurred in the comparison group is less meaningful in terms of personalization than the type of learning achieved by the target group. However, at this stage, this should be considered a tentative hypothesis; further research is needed to test and compare the two conditions with more data and in more detail.

Two differences are apparent based on analyzing the writing processes of the two observed groups based on Extracts 1 and 2: (1) the students’ (lack of) certainty about how to progress, and (2) how the ideas developed in the group discussions were incorporated in the essays.

First, the students in the target group appeared more confident than those in the comparison group when discussing what ideas to include in their essays and when using personal stories to anchor their knowledge. However, the variety of ideas surfacing in these students’ discussions appeared not to be entirely beneficial; this is revealed, for example, when Jane discovered that her text was not well organized (“But I think the structure of my text is bad,” line 12 in Extract 1). However, the teacher did not address this issue. In a similar vein, the students in the comparison group was uncertain about their knowledge of the assigned topic and if they were progressing, as revealed by Mike in lines 21 and 23 of Extract 2: “OK. But I feel if I want a better grade, there is something more I have to change.” “And I don’t know what it is. That’s kind of annoying.” However, the two boys showed more confidence when it came to applying knowledge of essay organization. They suggested different strategies, including cutting down text, building up a complex argument by splitting a paragraph, shortening a long passage or expanding another one.

Second, the ideas prompted by EssayCritic in Extract 1 were worked with twice in the observed target group (Carol and Jane): first when discussed in their group (both tell personal stories) to trigger and elaborate common ideas and later when incorporated in their essays as individual writing efforts. In Jane’s essay, the ideas discussed in her group were incorporated as one phrase whereas Carol wrote two sentences. The subthemes did not surface explicitly in the discussions of the comparison group (after reading Josh’s essay, Mike says in line 9, “I have looked at yours and it’s very good, there is very little to correct”). The two students struggled to generate feedback for each other in terms of subject matter content, even though they had access to the assessment rubric and the list of subthemes on a sheet of paper. They seemed to take on the role of the teacher when trying to create feedback for each other, possibly imagining what the teacher would have said in that situation. However, they found this process to be out of their league, as revealed by Josh in line 6 of Extract 2: “I really hate giving feedback because it is so difficult.” The students’ lack of competency in this area might have contributed to their uncertainty. Despite this, Josh and Mike received on average the same grade as Carol and Jane on the final essay (see Table 2), and some of the sentences they wrote can be traced back to their discussions.

The comparison group’s understanding of how to structure an essay might have been one factor to explain their final grades. Other possible explanations are as follows: (1) the comparison group had a better starting point, i.e. their version 1 essays received higher grades than those of the target group (3.5 vs. 2.7), (2) the teacher intervened twice as often in the comparison class than in the target class (13 times vs. 6), (3) it is possible that the two boys were not able to verbalize their ideas but were able to express them in writing because the ideas did not originate through collaboration, and (4) the covered subthemes feedback from EssayCritic (Figure 1: left) did not provide explicit guidance for essay reorganization for the target group students.

Writing and discussing to learn through the use of EssayCritic is supported by the design-critiquing framework (Fischer et al., 1991; Robbins & Redmiles, 1998; Mørch et al., 2005), which suggests the use of a cyclic process of action (writing) and reflection (conversing and thinking aloud) to explain how feedback is used in practice. When the action-reflection loop is sufficiently tight, the learners may effectively use the feedback generated, as it is fresh in mind and relevant to their task. This might be the factor that allows feedback to serve as a bridge
between prior knowledge (personal experiences in our case) and teacher set goals (desired knowledge defined by the assignment).

A possible weakness of the approach to learning analytics for foreign language learning we have shown in this study can be traced back to the eleven subthemes that were programmed in the EssayCritic learning algorithm (Appendix B). The choice of subthemes might have been constrained by the content of the textbook used by the school used in this study and the pedagogical preferences of the teachers. Further research ought to investigate possibilities for distinguishing between sufficient and insufficient number of subthemes to extensively cover a particular topic.

Summary and conclusions

We have developed an English essay writing aid through several iterations, applying different algorithms for comparing a student essay with good examples represented by a conceptual model. The latest (third) version uses a decision tree supervised learning algorithm, a carefully constructed concept tree of the topic (“English as global language”), and a carefully prepared knowledge base (teachers and researcher). This novel approach has a distinct advantage; even a complex or abstract subtheme can be represented by some easily evaluated logic rules, which was not possible with the algorithms used in the previous versions of EssayCritic. Our design and analysis efforts were informed by a theoretical framework (DCF) that takes into account individual (writing) and collaborative (discussion) activities around shared artifacts (essays), iterative design (versioning), and feedback. We compared two conditions of feedback provision (computer-generated and peer assessment), involving five classes at an upper secondary school in Norway (three classes to provide a knowledge base and two classes for direct involvement in the experiment). We used a mixed methods approach to gather and analyze the data: to compare quantitatively the outcomes according to grades and subthemes and another for “zooming” into the qualitative differences in the students’ writing process with examples. The quantitative data shows that there was no significant difference between the final grades in the two classes, but there was a significant difference in the number of subthemes. The qualitative data shows that the target group put more effort into writing essays rich in content, i.e., including many ideas, some of which were triggered by feedback from EssayCritic and the students’ personal experiences. The comparison group found it difficult to provide feedback on content and compensated by focusing on essay organization of fewer ideas.

Further research that would build on our findings could include, but is not limited to, the following:

- Automating feedback that would assist students in organizing their essays
- Computational support for analyzing deliberations (conversing) in addition to artifacts (essays) for a full support of the DCF. This could be achieved via a discussion forum and/or by capturing conversations
- Exploring whether the students who used the feedback from EssayCritic are able to transfer the processes (rather than content) to other writing assignments (such as feedback giving skills)
- Comparing the effects of discussion groups with individual work when using EssayCritic to identify other processes and mechanisms students use when writing to learn in groups other than collaboration
- Reanalyze the data material for multiple triggers of deliberations (external and internal), including deliberations triggered by EssayCritic, peers and self-triggered deliberations
- Exploring the tentative hypothesis that new knowledge is connected to prior knowledge that can be externalized (personal stories; verbalized experience). What methods can help to reveal the “gap closing” activity of prior and desired knowledge, and what role should the computer play in offloading (e.g., capturing traces of past activity for triggering memory, tests to determine proficiency according to levels, and so forth)?
- Study how the ideas developed by students change over time in response to feedback and revision, and how the ideas travel back and forth between written work and small group conversations.

Acknowledgements

The research design and data collection were carried out in the Ark&App project funded by the Ministry of Education of Norway. We thank Øystein Gilje, Sten Ludvigsen, and Anne Edwards for their constructive comments on an earlier version of this article.
References


Appendix A

Illustration of a subtheme, the decomposed simpler concepts, and the rule learned

Example subtheme: **English as a dominant language on the Internet and computers.** Table 4 shows the simpler concepts and the corresponding words or phrases (created manually by teachers and researchers), and Table 5 is the rule learnt by computer for identifying the subtheme in student essays.

**Table 4. The simpler concepts of a subtheme and the corresponding words or phrases**

<table>
<thead>
<tr>
<th>Simpler concepts</th>
<th>Words or phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>English</td>
</tr>
<tr>
<td>Communication</td>
<td>communicate; communication; talk; talked; discuss; speak; spoken;</td>
</tr>
<tr>
<td>IT areas</td>
<td>computer; programming; web; internet; www; information; cyberspace; website; webpage; YouTube; Netflix</td>
</tr>
<tr>
<td>Pronoun*</td>
<td>it</td>
</tr>
<tr>
<td>Example*</td>
<td>example; instance; e.g.,</td>
</tr>
</tbody>
</table>

**Table 5. The logic rule learned for the example subtheme**

<table>
<thead>
<tr>
<th>Subtheme</th>
<th>Logic rule composed of simpler concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>English as a dominant language on the Internet and computers</td>
<td>(“English” AND “IT areas”) OR (“Pronoun” AND “Communication” AND “IT areas”) OR (“Example” AND “IT areas”)</td>
</tr>
</tbody>
</table>

Appendix B

List of subthemes of English as a global language

1. Historical reasons for the spread of English language
2. Impact of English on other languages
3. English as Lingua Franca (travelling, international policies and diplomatic negotiations)
4. English as a dominant language on the Internet and computers
5. English as a language of science and technological innovations
6. English grammar and borrowed words
7. Advantages of English in University studies abroad
8. The US influence on the world (economy, film industry)
9. Knowledge of English helps in finding high quality jobs
10. English as an International standard for communication for pilots and air traffic controllers
11. English vs. other official UN languages

Appendix C

Transcript notation

- [ ] Text in square brackets represents clarifying information
- = Indicates the break and subsequent continuation of a single utterance
- ? Rising intonation
- : Indicates prolongation of a sound
- () Short pause in the speech
- [...] Utterances removed from the original dialog
- - Single dash in the middle of a word denotes that the speaker interrupts herself
- -- Double dash at the end of an utterance indicates that the speaker’s utterance is incomplete
- ((Italics)) Annotation of non-verbal activity
Assessing the Language of Chat for Teamwork Dialogue

Antonette Shibani¹, ³, Elizabeth Koh*, Vivian Lai² and Kyong Jin Shim²

¹National Institute of Education, Nanyang Technological University, Singapore // ²Singapore Management University, Singapore // ³University of Technology Sydney, Australia // antonette.shibani@gmail.com // elizabeth.koh@nie.edu.sg // vivian.lai.2011@sis.smu.edu.sg // kjshim@smu.edu.sg

*Corresponding author

ABSTRACT

In technology enhanced language learning, many pedagogical activities involve students in online discussion such as synchronous chat, in order to help them practice their language skills. Besides developing the language competency of students, it is also crucial to nurture their teamwork competencies for today’s global and complex environment. Language communication is an important glue of teamwork. In order to assess the language of chat for teamwork dimensions, several text mining methods are possible. However, difficulties arise such as pre-processing being a black box and classification approaches and algorithms being dependent on the context. To address these issues, the study will evaluate and explain pre-processing and classification methods used to analyze teamwork dialogue from a dataset of chat data. Analytics methods evaluated in this study provide a direction for assessing the language of chat for teamwork dialogue and can help extend the work of technology enhanced language learning to not only focus on academic competency, but on the communication aspect too.

Keywords

Teamwork, Pre-processing, Supervised machine learning, Text mining, Learning analytics

Introduction

In technology enhanced language learning, many pedagogical activities involve students in online discussion such as synchronous chat, in order to help them practice their language skills (e.g., Hedayati & Foomani, 2015). Such online team chat increases students’ participation and evenly distributes their participation (Dennis & Garfield, 2003). In today’s global and complex environment, besides developing the language competency of students, it is also important to nurture their teamwork and collaboration competencies (Rychen & Salganik, 2003). Moreover, language communication is an important glue in teamwork that helps build the team together and propel it forward in various tasks (Baker et al., 2005). Thus, beyond evaluating the academic competency of the language, it is meaningful to examine the teamwork discourse of students. Providing an assessment of teamwork dialogue helps students gain a better awareness of their teamwork competency and become better team players (Erkens & Janssen, 2008; Koh et al., 2014). Advancements in technology like learning analytics have allowed such assessments to be made more automated.

To develop such automated assessment systems, training models using human evaluation (involving computational linguistics and text classification) have the highest potential for efficiency and scalability. These models use computers to code the data instead of humans, which is faster; they can also be developed to examine different texts which allows for scalability (Erkens & Janssen, 2008; Rosé et al., 2008). In other words, language assessment can be made more efficient, and this facilitates more immediate feedback for students and teachers which can help improve students’ learning (Anjewierden et al., 2007). However, a major challenge in automatically assessing online chat discourse is that chat texts have many irregularities in structure, short lengths and contextual complexities. This makes the identification of codes (such as teamwork dimensions) more difficult. Past research suggests that pre-processing the text to organize and take into account desired features would be helpful for analysis. Despite the importance of pre-processing, the steps in pre-processing tend to be vague. There seems to be a black box of pre-processing.

Another problem is that there are varying approaches to train models using human evaluation such as natural language processing and supervised machine learning. This area of analysis has not been widely documented and techniques also depend on the nature of the coding scheme and the purpose of classification. Furthermore, there are many machine learning algorithms that could be used. Finding out the most effective method would be the key to provide an automatic analysis of the language of chat, which serves as formative assessment for students.

Therefore, the focus of the paper is on the methods of using learning analytics for assessing online chat discourse, in particular, to measure the dimensions of teamwork. The research questions are:

- How can text be effectively pre-processed to assess the language of chat for teamwork dialogue?
- What approach works best to assess the language of chat for teamwork dialogue?
- Which algorithms are the most effective in classifying teamwork dimensions?
This study is part of a larger research project exploring the 21st century competency of teamwork in technology enhanced learning. Based on previous literature and pilot studies, six dimensions of teamwork were conceptualized (Koh et al., 2014). A coding scheme for these dimensions was developed and chat log data was manually coded. Chat log data of the study was obtained from 14 year old students who were collaborating on an online collaborative problem-solving activity as part of their project work curriculum. Besides providing students with the opportunity to engage in teamwork, the activity was designed for English communication, which also helps students practice their language skills. Baker et al. (2005) conceptualizes that communication “is the glue that holds the team together” (p. 240). In that sense, while there are specific teamwork dimensions, communication is an important part of the team as it joins the team together. In other words, while we measure teamwork dimensions, we are also indirectly measuring their language communication competency. Thus, it is possible to infer that if students have high teamwork competency dimensions, their communication skills will be high.

This paper will contribute to the research on learning analytics methodology for language learning. It attempts to open up the black box of pre-processing, revealing hidden steps in many learning analytic studies. The paper will also reveal approaches and algorithms most effective in identifying teamwork dimensions, which will be relevant to educational data mining and learning analytics researchers. Lastly, the empirical implementation and results provide an applied context for using learning analytics, and offer some insights for educational practitioners in the potential uses of learning analytics.

The paper begins with a brief literature review of the methodology involved in text mining in online chat. This is followed by further details of the background of the project and the dataset. Next, we elaborate on the methods of the study and look at both pre-processing and classification approaches. Subsequently, we test out our methods and discuss the results. The conclusion section deliberates on the practical and theoretical implications of the study.

**Literature review**

Online collaboration platforms such as Moodle, HipChat and Slack provide insights into individual and organization-level communication and collaboration behavior (Anders, 2016), help identify indicators of academic performance in students’ online forum participation (Romero et al., 2013); they also identify patterns of student interactions and participation (He, 2013). Moreover, automated text mining analysis allows students to make better sense of the learning process during their collaboration, and possibly provide support and remediation (Anjewierden et al., 2007; He, 2013). This is an application of “Learning Analytics” which is defined as the “measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Siemens & Gašević, 2012, p.1).

Text mining traditionally analyzes long text such as essays and newspaper articles. However, chat text length is short and has irregularities in structure and contextual complexities. One of the key decisions for learning analytics researchers is whether to pre-process such text. Pre-processing is a core task in text mining to clean up the raw data and structure it for analysis. In structured text, the general pre-processing steps include: removing punctuations, trimming white spaces, performing stemming and POS tagging, and removing stop-words. Pre-processing can help clarify the meaning of text, but is also highly dependent on the data and the context. Steps of pre-processing are not widely documented in learning analytics. Many studies only do a basic pre-processing of text data since the usefulness of contextual pre-processing has not been studied in depth. Some studies choose not to perform pre-processing, as the original features of the text, e.g., punctuations, can be useful to derive meaning (Villatoro-Tello et al., 2012). On the other hand, a study on automatic topic detection in chats used sessionalization and extraction of features from icon text and URLs for pre-processing (Dong et al., 2006). As aforementioned, chat texts are rather different from conventional text, and could require more or different ways of pre-processing.

Several systems have been developed to take care of these issues in chat text pre-processing using the process of normalization (Clark & Araki, 2013; Han et al., 2012; Rosé et al., 2008). In addition to the misspelled words, normalization also corrects common chat language comprising out of vocabulary words (OOV) such as acronyms, abbreviations and emoticons. These systems are either dictionary-based or algorithm-based. A dictionary based approach was shown to outperform sophisticated algorithm approaches in terms of F score and word error rate by generating possible normalization pairs (Han et al., 2012). In a system called Casual English Conversion System (CECS), eight categories of irregularities such as Shortform (abbreviation), Shortform
(acronym), Typing error/misspelling, Punctuation error/omission, Non-dictionary slang, Cultural reference/in-group meme, Wordplay/intentional misspelling, and Omission of vocabulary were captured (Clark & Araki, 2013). This system preserves the phrase structure in texts and makes it more context-aware, hence useful for our work.

Pre-processing is typically followed by classification, where a discrete category is predicted based on the training data. Most approaches use human coding as the standard to check if the automated coding classifies instances similar to the standard (Anjewierden et al., 2007; Erkens & Janssen, 2008; Rosé et al., 2008). One application of classification approaches is in the identification of “topic” when there are multiple participants contributing text into a single common chat room.

There are several classification approaches depending on the nature of the coding scheme and the purpose of classification. One such approach is a rule-based classification based on the textual features and patterns identified for each dimension (Aggarwal & Zhai, 2012). Such classifiers are written using Natural Language Processing (NLP) rules for automating qualitative data analysis. Crowston et al. (2012) used a symbolic approach by looking for evidence of specific behavioral patterns and writing rules for automating content analysis. Different levels of analysis from the lowest phonological to the highest pragmatic linguistic approaches can be done. They do not require large datasets like statistical approaches, but are not easily transportable to other domains. The results of automated coding were compared with manual coding and a good performance on a number of codes was seen in terms of precision and recall measures. The study prioritized higher recall over precision, so that it can act as a support system for human coders to revisit and change codes if necessary. In another study, Erkens and Janssen (2008) developed an automatic coding procedure for 29 dialogue acts using the Multiple Episode Protocol Analysis (MEPA) computer program consisting of if-then rules using discourse markers and cue phrases for pattern matching. Another approach is supervised machine learning, where the aim is to classify messages according to a pre-determined set of categories. Algorithms such as Naïve Bayes, Support Vector Machines and k-Nearest Neighbor are used in machine learning to automatically categorize chat messages (Anjewierden et al., 2007; Rosé et al., 2008) in an educational setting to analyze student chat conversations. In line with the aforementioned studies, the current study will use and evaluate pre-processing and classification techniques to identify teamwork dimensions from students’ chat data.

Research context, corpus and coding scheme

This research is part of a larger study on teamwork. A teamwork competency awareness program was designed for 14 year old students in a Singapore secondary school, for them to gain an awareness of their teamwork competency. During one section of the program, student teams of 3 or 4 used an online group chat to complete a problem-solving activity. The first task was an ice-breaker activity, while the second was a dilemma task. Students had about 45 minutes to complete the tasks using computers in school during their class time. Students were seated away from their team members to reduce face-to-face communication and encourage online communication. Instructions were provided to the students through the chat in real time by the Chat Administrator, who is a researcher. As one of the aims of the study, we hoped to help students gain an awareness of their teamwork competency through multiple measures. One measure was the assessment of teamwork dimensions of the online chat dialogue.

The six dimensions of teamwork competency measured are: coordination (COD) - organizing team activities to complete a task on time; mutual performance monitoring (MPM) - tracking the performance of team members; team decision making (TDM) - integrating information, selecting the best solution, and evaluating the consequences in a team; constructive conflict (CSC) - dealing with differences in interpretation between team members through discussion and clarification; team emotional support (TES) - supporting team members emotionally and psychologically; and, team commitment (TCM) - identifying with and being involved in team goals (See Koh et al., 2014 and Koh et al., 2016 for further descriptions). A sample of the coding scheme is provided in the Appendix.

A total of 272 students participated in the online activity in 76 teams. This resulted in 19762 raw chat messages, inclusive of spam lines. This formed the corpus of the study. Two coders annotated 7 teams’ data with a Cohen’s Kappa > 0.65. They then proceeded to code the rest of the teams individually. For this analysis, data from 34 teams coded manually was taken for automated analysis. A sub-section of the dataset was selected for this analysis due to practical reasons as the approach of NLP rules requires substantial time for rule creation. This dataset had 9783 lines in total which was split into test and training sets. The minimum was 71, the maximum was 487, and the average was 287 lines per team. The training set consisted of 5705 lines and the test set, 4078
lines. The training set was selected using maximum variation sampling to contain diverse data based on participation, class and team composition. It represented data from high and low participating teams in terms of chat lines, teams from all 7 classes and both 3 and 4 member teams, for the results to be more generalizable to new data.

Methods and methodology

The different approaches and methods developed are described in this section.

Pre-processing

With regard to text pre-processing, two types of text were created: (1) non-pre-processed text, termed “base,” and (2) pre-processed text.

For the base text, the steps comprised:

- Removing unicode text from raw text - to overcome Python unicode decode error
- Situation coding and spam filter - this was a manual coding performed to categorize messages into situations and spam/no code (Shibani et al., 2015).

The chat lines were grouped by topics for eight situations starting from Introduction (ST1) to Team Dismissal (ST8). These situations reduce the ambiguity in the context of data and help in writing rules for classification. The messages that need not be coded for teamwork dimensions (mostly spam) were also coded as “no code” (nc) manually while performing situation coding. With our reasonable amount of data, it was easier to do manual spam detection since developing a context-sensitive spam detection system was out of scope.

While some of the spam lines seem to not interfere with the text classification, some do, since they contain keywords from teamwork dimensions. Certain dimensions like COD and TES follow the same structure and hence it becomes necessary to distinguish the chat lines from nc lines that need not be coded. Also, these lines when given as input decrease the performance of machine learning algorithms, since the machine learning systems will extract features from this text and also predict categories for them. The rules defined to isolate nc lines can be found in the Appendix.

For the pre-processed text, the main idea was to simplify the text for the classifier to learn the features easily. For example, there can be just one feature {{Name}} in a feature set, instead of having the classifier to learn 100 names from the text as features. We hope that this will help in grouping similar features together to build a better predicting classifier, which is also more memory and time efficient.

The pre-processing mechanism covers several features of chat data and preserves useful nuances in text. Figure 1 shows the different steps of pre-processing that were performed for normalizing chat data. We next describe each of these steps.

Emoticons and punctuation tagging

As emoticons and several punctuation marks were important to our coding scheme, we converted them to indicators (tags). A list of emoticons were compiled from online sources and our own data, and grouped into categories (See Appendix). Other punctuation marks that were not relevant to the coding scheme were removed.

Chat abbreviation expansion

Chat language contains many abbreviations in the form of short forms and acronyms that have to be expanded. Short forms are shorter representations of a word by omitting or replacing few characters e.g., grp → group. Acronyms are formed from the first character of each word e.g., lol → laughing out loud. These words may be omitted by keyword search due to incorrect match if not expanded. We created a dictionary of these words from urban dictionaries (see http://www.urbandictionary.com/) and acronym dictionaries (see http://www.singlishdictionary.com/) to replace abbreviations by expansions. Relevant words were also added
from the CECS database for more comprehensive coverage for other corpus data (Clark et al., 2013). It is to be noted that all chat abbreviations cannot be replaced by an expansion, as some meanings vary in different contexts, e.g., “btw” may refer to “between” or “by the way.” Contractions and expansions in English are also taken care of in this step, e.g., can’t → cannot. Words specific to our context were also added as they occurred frequently in the chat, e.g., ans → answer.

Figure 1. Steps in pre-processing chat data

Local terms replacement

In addition to the chat terms, there were also local terms found in the corpus, whose meanings cannot be found in an English dictionary. These Singlish terms derived from the local languages are extensively used in chat by students. We constructed a local dictionary with Singlish terms and their English equivalents for use in our system. The words in this dictionary were gathered from websites which provide meanings for Singlish terms and also from our own corpus. The local terms were replaced with their equivalent English terms using this dictionary, e.g., ah ma → grandma, cher → tea cher, kampung → home.

Named entity recognition

The reference of names could not be ignored in our context as it indicates teamwork dimensions in different circumstances according to our coding scheme e.g., in COD for discussing who is in the team (“John’s in our team”). Named entity recognition identifies names and labels the different elements of text into predefined categories like person, organization, numbers etc.

We used the Stanford NER caseless class model trained for CoNLL and MUC data that tags three entities: Person, Location, Organization, Misc (Finkel et al., 2005). The NER is part of the Stanford CoreNLP that contains a list of open source tools for natural language analysis. The caseless model that ignores capitalization was selected as most of our chat data did not follow proper capitalization of names. We ran a manual check to remove the wrongly identified Person tags. The pre-processing script then tagged the names found in the chat as {{Name}}. Training our own classifier for local names could be part of future work using more data, as the existing Stanford NER models cannot be extended.

Spelling correction

Spell checking is performed as the next step of the pre-processing pipeline on all word tokens excluding the already tagged ones from the previous steps. Upon this transformation, all words were tagged with a {{S}} prefix to indicate that they were changed. The engine is written in Python using PyEnchant package (see https://pypi.python.org/pypi/pyenchant/) and works in three stages as shown in Figure 2.
The goal of our classification task was to classify chat lines into zero, one or more teamwork dimensions. The main issue for automatic classification is that we require a single line to be coded for multiple dimensions (multi-label classification), whereas most algorithms classify only one category to an instance (binomial or multi-class classification). In multi-class classification, even if there are six classes of teamwork dimensions, they are mutually exclusive and each chat line can be categorized as only one of the dimensions. In multi-label classification, each instance can be assigned multiple labels, like a chat line indicating both COD and TES can be assigned both. In our current implementation, we have classified each dimension separately as 0 or 1 (binomial classification) for a line and then combined all dimensions together to get the final predicted output.

The two approaches below were implemented for classification:

- NLP rule-based classification
- Supervised machine learning

We shall discuss the implementation of each method in detail and then evaluate their performances using our dataset.

NLP rule-based classification

Our first approach was to do a rule-based classification based on the textual features and patterns we identified for each dimension (Aggarwal & Zhai, 2012). The unit of analysis is a chat line. Text is represented as a string with a sequence of words in order to preserve useful information like the sentence patterns and phrase structures in the chat lines. To maintain the contextual link between the preceding line and the next line and sentence cohesion, the order of chat lines was also preserved. A search is performed on the identified features in the text and the resulting matching text are coded for teamwork dimensions. This system was developed in R. The rules are specific to each dimension, but the predicted labels are combined to obtain the final coded output. Table 1 displays a sample.

The indicative terms dictionary consists of keywords that are indicative of teamwork dimensions and forms the basis of the classification system. The dictionary may contain individual words or phrases depending on the context. The following steps are used to write rules based on the indicative terms (Shibani et al., 2015).

Existence: Some keywords from the indicative terms dictionary are strongly indicative of the teamwork dimensions. This means that the occurrence of such words or phrases demonstrates that the chat line contains a particular teamwork dimension. For example, a positive emoticon will always indicate TES and hence it can be detected by its existence. Similarly task-specific words like factory, pollution etc. indicate TDM. For each dimension, an input string is searched for such terms from the indicative terms dictionary. If the term is present, the teamwork dimension is coded 1, else it is coded 0.

Frequency: The frequency of indicative terms can indicate the strength of a dimension in a chat occurrence. We do not measure the strength of a dimension in the current implementation and we only measure the presence, i.e., a line containing 5 keywords for TDM and another line containing a single keyword for TDM are both coded the same 1 for the dimension. In systems that measure this variation, such rules for frequency can be written. However, we used the n-gram tokenizer from Weka (see http://weka.sourceforge.net/doc.dev/weka/core/tokenizers/NGramTokenizer.html) to find the frequent bigrams.
and trigrams in addition to single words that occur in the chat lines and determined if they contributed to teamwork dimensions. These frequent words identified from the chat text were added to the indicative terms dictionary rather than manually looking for keywords, thus reducing human effort.

<table>
<thead>
<tr>
<th>Name</th>
<th>Message</th>
<th>Situation</th>
<th>COD</th>
<th>MPM</th>
<th>TDM</th>
<th>CSC</th>
<th>TES</th>
<th>TCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Hellooooo</td>
<td>ST1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>Hi :D</td>
<td>ST1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>WHO ELSE AH?</td>
<td>ST1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>LET S DO THIS</td>
<td>ST2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Admin</td>
<td>a) Describe your ideal teacher</td>
<td>ST2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>DESCRIBE YOUR IDEAL TEACHER :)</td>
<td>ST2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>bob describe cher sia</td>
<td>ST3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>understanding lor</td>
<td>ST3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>DISCRIBE CHER????</td>
<td>ST3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>a) understanding b) respectful ,dont give much homework</td>
<td>ST4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>okay done</td>
<td>ST4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>:like:</td>
<td>ST4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Sample coding for teamwork dimensions in a chat log

**Proximity:** Proximity can be used to identify the context of words in utterances. For example a salutation followed by name can be identified as a teacher’s name rather than a student’s name (“Mr Wong is cool, agree Amy?”). Rules based on the proximity of keywords were identified from the corpus. While the rules involving keywords within the same utterance in a particular proximity are relatively easier to implement, the rules involving threaded utterances are more complex. An example is finding a word corrected for spelling by a student from the previous line (Line 56: Write about your ideal teacher, Line 58:*ideal). This involves bringing the previous and next lines into context and requires complex implementation.

**Weightage:** There are a few keywords that indicate one dimension but also occur in other contexts. Such words were removed or added based on the error analysis of confusion matrix, obtained by comparing human and automated coding, giving weightage for one dimension over another. We aimed to reduce false positives where the line is coded for a dimension when it is not indicative of that dimension (0-1, where 0 denotes manual coding and 1 denotes automated coding). For example, if a keyword for TDM increased the 0-1 pairs for COD with no considerable increase in 1-1 pairs of TDM, it was removed.

**Complex patterns:** More complex patterns involving coded situations are useful in identifying the context of text. For example, task related keywords should occur in the designated situations like ST3 and ST6 where the task discussion occurs, and greetings occur in ST1. A name occurring in ST1 is part of a rule for COD, hence a related rule can be written to distinguish it from MPM (ST1 + name in text represents COD). Also, typo correction by self or other members can be used to distinguish MPM or TES, with the help of complex rules.

Although the above approach looked fine for the data we analyzed after initial coding, there are inherent disadvantages in this method. The key issues are as follows:

**Generalizability to new data:** Unless the future data is very similar to the current data, the classifier will not generalize, which is the case in most real-world scenarios. The students from a different school can have a completely different style of writing. This leads to the case where the teamwork dimensions are not captured by the classifier, because the words used are different from the previously identified words and patterns that the classifier was trained and tested on. An ideal classifier should be able to generalize to new data by learning features that can be applied to new data without high variance.

**Labor intensive rule writing:** The process of identifying rules from the data and implementing them in a classifier is a time consuming and laborious process. If this process has to be repeated for all the new data, it increases the workload. Also, if there are changes in the coding scheme, the rules may have to be modified to fit the new coding scheme. Such manual feature extraction process makes the effort demanding and strenuous.
**Context-sensitive rules:** The NLP rules which are used are human-context sensitive and the rules can be written in different ways by different researchers, giving inconsistent results. It is very difficult for a researcher to reproduce/replicate another researcher’s rules just by looking at the instructions and data, since they may identify different patterns. Also, there is no common standard for rules or guidelines for the number of rules. Another related issue would be overfitting, when too many rules are added such that it reduces the classifier’s ability to generalize.

**Supervised machine learning**

Our second approach was to implement a classifier using machine learning algorithms. Since the human coded labels were available for training the classifier, supervised machine learning was used. Supervised learning algorithms learn features from the labelled instances and use them to predict the category for new unlabeled instances based on the likelihood suggested by a training set (Kotsiantis et al., 2007). The same training and test datasets from NLP rules classifier were used for analysis so that both the methods can be compared.

The advantage of this method is that the classifier can extract features from the training data automatically, so there is minimal human effort in identifying features. It also makes use of existing implementations of the different machine learning algorithms, so there is no need for a new development of a system except for tuning the required parameters to obtain optimal classifiers.

The scikit-learn Python package (see http://scikit-learn.org/stable/) was used for this implementation as it allowed customizable scripting and the use of different algorithmic implementations without requiring expert technicalities. We implemented the following six commonly used algorithms for our analysis and comparison (Aggarwal & Zhai, 2012, 2010, Kotsiantis et al., 2007):

- Decision Tree (DT)
- K-Nearest Neighbors (KN)
- Logistic Regression (LR)
- Naïve Bayes (NB)
- Single-layer Perceptron (PE)
- Support Vector Machine (SV)

Our machine learning classifiers used the textual chat message as input to extract features. Both the base text and the pre-processed text of the messages were used as inputs for comparison. The spam and “nc” messages were removed from the training and test data so that the classifier learns better features as explained in the earlier section. To represent the text document as a vector for machine learning, the documents are typically stored in terms of the word features and its occurrences in the document. For our data, the messages were vectorized based on counts. In large corpus, the tf-idf would have worked better since it captures the words which are more interesting in the given documents by considering both the term frequency and the inverse document frequency. The CountVectorizer from scikit-learn was used to tokenize and extract both unigrams and bigrams. Using both unigrams and bigrams as features gave better results than using unigrams only. This was probably because our data had useful features that occurred together as bigrams. Feature selection based on statistical tests did not improve the classification results since it removed useful features from our small dataset, thereby decreasing performance, so we have currently used all the extracted features.

Future work will focus on implementing a multi-label algorithm, where all six dimensions can be classified by the same classifier rather than combining multiple binary classifiers. This will overcome the problem of finding dependencies among the dimensions. The disadvantage of this traditional bag-of-words approach is that it does not consider the semantic relations between words, so it is difficult to improve the accuracy of these classification algorithms unless more features are identified from a bigger dataset.

**Results and discussion**

To address research questions 1 and 2, four common metrics were calculated: Cohen’s Kappa, F score (sometimes called F measure), precision and recall (Powers, 2011). Accuracy was not calculated due to certain limitations (e.g., skewed classes with unbalanced data). Human coded data was considered the gold standard, and compared with machine coded data. Cohen’s Kappa (see http://www.pmean.com/definitions/kappa.htm) is a commonly used statistical measure to calculate inter-rater reliability between two coders, controlling agreement by chance. Minimum acceptable Cohen’s Kappa values range from 0.4-0.6. Precision is defined as the ratio of
the number of correctly coded cases by the machine learning model to the total number of cases coded by the model. Recall is defined as the ratio of the number of correctly coded cases by the model to the total number of correctly coded cases specified by standard human coding. A model has high precision when most of its cases are coded accurately, which however, reduces recall since it does not code many cases. To maintain a balance between these two measures, the F score calculates the harmonic mean of the precision and recall to create a single measure for model effectiveness. Although the preferred values for a good model depends on the context of data, 0.6 and above is considered an acceptable standard for reporting these measures in educational data mining.

These metrics were calculated for each of the six teamwork competency dimensions. A comparison between the type of text, namely base and pre-processed text was performed for each of the two classification approaches, NLP rule-based classification (NLP rules) and supervised machine learning (Sup ML). Table 2 reports the performance metrics.

<table>
<thead>
<tr>
<th></th>
<th>COD</th>
<th>MPM</th>
<th>TDM</th>
<th>CSC</th>
<th>TES</th>
<th>TCM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohen’s Kappa</td>
<td>0.35</td>
<td>0.53</td>
<td>0.18</td>
<td>0.36</td>
<td>0.69</td>
<td>0.72</td>
</tr>
<tr>
<td>F score</td>
<td>0.46</td>
<td>0.63</td>
<td>0.21</td>
<td>0.42</td>
<td>0.80</td>
<td>0.83</td>
</tr>
<tr>
<td>Precision</td>
<td>0.59</td>
<td>0.81</td>
<td>0.73</td>
<td>0.83</td>
<td>0.85</td>
<td>0.92</td>
</tr>
<tr>
<td>Recall</td>
<td>0.38</td>
<td>0.62</td>
<td>0.12</td>
<td>0.33</td>
<td>0.75</td>
<td>0.83</td>
</tr>
</tbody>
</table>

| **Pre-processed** |     |     |     |     |     |     |
| Cohen’s Kappa | 0.29 | 0.56 | 0.29 | 0.59 | 0.69 | 0.74 |
| F score       | 0.46 | 0.64 | 0.33 | 0.64 | 0.80 | 0.84 |
| Precision     | 0.43 | 0.79 | 0.78 | 0.78 | 0.84 | 0.96 |
| Recall        | 0.51 | 0.62 | 0.21 | 0.59 | 0.77 | 0.84 |

Note: Results for the highest performing classifier algorithms are recorded, with the algorithm listed in the brackets. In bold are the metrics in which pre-processed text was larger than base text. In italics are the metrics in which Sup ML was larger than NLP rules.

![Figure 3](image_url). F score performance of text type and classification approach for the 6 teamwork dimensions
Figure 3 provides a graphical representation of the F score for two comparisons of text type and classification approach. As can be seen, pre-processed text performs better than the base text on most teamwork dimensions. Examining F score results of NLP rules, pre-processed text outperformed base text in 3 out of the 6 dimensions, tying in 2 dimensions, and decreasing by 0.01 for CSC. More obviously, F score results of Sup ML showed better performance of pre-processed text compared to base text for 5 of the 6 dimensions, and tying with one dimension, TCM. This suggests that pre-processing text is better than non-pre-processed text with regard to our dataset and features.

As for classification approach, it is clear from Table 2 that Sup ML outperforms NLP rules in most of the metrics. The only exception was in CSC where the NLP rules were written to mimic human coding by using TDM as a rule, since a part of TDM lines also fall under CSC. In machine learning, since the dimensions were independent, there was no link between TDM and CSC. Since the classification of CSC was only based on the word features in machine learning, it performed lower. This suggests that Sup ML is a better method for the dataset and features compared to NLP rules in general. It is also better to go with Sup ML for scalability reasons, as it is able to learn features for new and larger datasets. However, one limitation of our current manual coding scheme is that it takes into account not only words, but turn-taking and member roles. Also, there is need for a larger dataset as well as other datasets to test out the performance of the approach. Further work is required to refine the machine learning features too. Nevertheless, NLP rules are a good second choice as these rules make sense to the users and are rather intuitive, although the rules are not too scalable for other contexts and tasks.

![Figure 4. Performance metrics of all algorithms](image)

Next, we examine research question 3, which compares how the 6 different algorithms classify teamwork dimensions. For the 6 dimensions, it was possible that different classifiers performed better in different dimensions, which was also observed in our analysis. There was no single classifier that performed the best among all dimensions, though SV performed higher than others in general. Figure 4 reports the results of all the metrics of the 6 algorithms. As can be seen, SV has the highest F score among the other 5 algorithms on almost
all dimensions of teamwork. It has also hit at least a 0.6 performance benchmark. Although SV has lower precision, recall is higher. This suggests that SV predicts more instances giving more coverage of dimensions implying that it is more complete than exact. The SV edges the other algorithms especially in terms of MPM and TES. It can also be seen that KN is the least effective algorithm for our coding scheme and dataset. These results are similar to Altrabsheh et al. (2014) who found that the SV algorithm had the highest F score, precision and recall for learning sentiment from students’ feedback.

KN is often seen to be intolerant to noise in the data which means that the irrelevant features in text corrupt its output. The irrelevant features in our dataset could have been an important reason for KN to perform badly for our data. SV on the other hand, removes irrelevant features and performs better. SV is suited to tasks with relatively larger number of features than number of instances since it only selects fewer support vectors. PE is also seen to have an advantage when there are few relevant features among many features due to their superior time complexity. One reason for NB to work reasonably well in our dataset is that it only requires a small dataset to train the classifier by estimating its parameters. It classifies correctly when the category is more probable than others. The classifier also ignores rough estimates in its underlying probability model to give a good overall classifier.

Other factors like runtime and memory were not considered for performance measures since the feature set was small and hence the training and test times were negligible. We intend to use the identified teamwork dimensions from this method by converting the occurrences into a scale, to provide an aggregated measure so that users (such as students) know the degree to which they have displayed the different dimensions of teamwork. Thus, measuring these teamwork dimensions serve as a formative assessment of students’ competencies.

Implications and conclusion

In this paper, we have evaluated and explained the pre-processing and classification methods used for analyzing teamwork from chat data. The study has found that pre-processed text is better than base text and re-iterates the importance of pre-processing according to the context of application. The pre-processing system has uncovered the complexities in chat language to better prepare the data for classification. It reveals unconventional English patterns in the educational context which are not covered in the existing corpuses. The process of how to deal with these irregularities is explained in detail which opens up the black box of pre-processing. This workflow for pre-processing chat data is a novel contribution, which can be potentially applied to other free flowing online chat text for normalizing raw chat to meaningful data. While time-consuming, our study suggests that developing and implementing the pre-processing techniques is a worthwhile investment.

The two different methods for classification have also been investigated using our dataset. The machine learning method seems to outperform the NLP rules for classification in terms of performance results. Practical considerations such as time and effort are reduced in machine learning with better scope for scalability. In supervised machine learning, we have also explored the capabilities of the different learning algorithms by comparing their performances. In general, the Support Vector Machine algorithm produces better performance for most teamwork dimensions in our dataset. This aligns with several papers in text mining of the effectiveness of the SV algorithm (Altrabsheh et al., 2014; Dong et al., 2006).

While the focus of our paper was to explain different methods to analyze online chat discourse with respect to teamwork dimensions, it should be noted that the results are specific to our data and the methods need to be selected appropriately for any given problem according to the context. Many of our rules also need further refinement. As shown, the metrics of the approaches can be further improved. Work is in progress to increase the performance of the classification approach. This will allow the study to reliably and relatively accurately categorize teamwork dimensions in a chat log. Future work can also include more quantitative “use frequency data” for analysis such as trace logs and counts of participation to see if they contribute to dimensions of teamwork. In the current context, such counts were not included due to spam lines in the data and the focus was more on assessing the quality of chat text.

Our current contribution enables efficient measurement of 6 teamwork dimensions in a semi-automated manner which can be used in a broad range of language learning situations. These dimensions do not require specific team member expertise, which is ideal in many foreign language learning situations where homogenous teams of students with no particular expertise in the language are learning together. This is a first step towards generating an automatic assessment of the chat language.
Moreover, this approach maximizes the learning experience of students. The learning process of students in online chat is used more than just for communication and topical dialogues, but the process itself is analyzed and becomes an artefact and means of learning. It becomes a reflective tool for students and part of their formative assessment. The research hopes to visualize the results of the automatic assessment in a learning profile that helps students better understand their language use and behaviors. We envision a scenario where this learning profile is shown at key points in the course so that students can stop, reflect and be more aware of their discourse. This feedback will go towards helping students understand their dialogic strengths and weaknesses for further improvement.

This approach also supports teachers and could make them more effective in large classes. As the system is easily scalable, in bigger classes, chat language can be automatically assessed and feedback provided efficiently to the students. The teacher is then able to zoom in on students who are not performing well, and intervene accordingly. For instance, a teacher can group large class sizes into smaller groups of students and provide them with a chat topic. After students complete their chat activity, the system can measure the teamwork dimensions and subsequently display the dimension scores to students. This provides feedback and allows students to become more aware of their teamwork competency dimensions. Teachers can also view the dimension scores of students and speak to the students who do not perform well. Overall, this reduces the load of teachers from assessing, and allows them to focus more on helping students to learn. That said, the teachers’ role in intervening in a meaningful manner is still crucial. While this method and approach makes certain communication skills of students visible, teachers still need to scaffold and help students make sense of the information and suggest ways for improvement.

The learning analytics methods evaluated in this study provide a direction for assessing the language of chat for teamwork dialogue. The approach and system will be useful for both students and teachers in evaluating the students’ communication ability. It will help extend the work of technology enhanced language learning to not focus only on academic competency, but the communication aspect too as informed by teamwork dialogue. Ultimately, the study provides a means of assessing chat dialogue such that students are able to gain a better awareness of their teamwork competency which is an important end in itself.

Acknowledgements

This paper refers to data and analysis from the research projects OER62/12EK and OER09/15 EK, funded by the Education Research Funding Programme, National Institute of Education, Nanyang Technological University, Singapore. The first author has subsequently moved to another institution. The views expressed in this paper are the authors’ and do not necessarily represent the views of NIE.

References


Appendix

Table A1. Teamwork competency coding scheme examples (Koh et al., 2014; Koh et al., 2016)

<table>
<thead>
<tr>
<th>Teamwork dimension</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordination (COD)</td>
<td></td>
</tr>
<tr>
<td>Organize activities to complete task on time</td>
<td>“Faster!”</td>
</tr>
<tr>
<td>Ask team members who is in the team</td>
<td>“Who is in our group?”</td>
</tr>
<tr>
<td>Coordinate logistics of task</td>
<td>“Combine the answer”</td>
</tr>
<tr>
<td>Mutual Performance Monitoring (MPM)</td>
<td></td>
</tr>
<tr>
<td>Give clarifying feedback to help in the team’s performance</td>
<td>“Shall we go and search for ideas?”</td>
</tr>
<tr>
<td>Ask team members to contribute to the task</td>
<td>“What do you think?”</td>
</tr>
<tr>
<td>Steer conversation back to task</td>
<td>“Ok back to the discussion”</td>
</tr>
<tr>
<td>Team Decision Making (TDM)</td>
<td></td>
</tr>
<tr>
<td>Give ideas related to the task</td>
<td>“Close down the factory”</td>
</tr>
<tr>
<td>Ask any task-related question</td>
<td>“What is most important?”</td>
</tr>
<tr>
<td>Exchange information about the task</td>
<td>“The factory can be moved to another location.”</td>
</tr>
<tr>
<td>Constructive Conflict (CSC)</td>
<td></td>
</tr>
<tr>
<td>Explain and give reason for disagreement</td>
<td>“I don’t completely agree because …”</td>
</tr>
<tr>
<td>Add on to ideas</td>
<td>“Also, it is very expensive to move.”</td>
</tr>
<tr>
<td>Propose different ideas</td>
<td>“How about adding filters to the factory”</td>
</tr>
<tr>
<td>Team Emotional Support (TES)</td>
<td></td>
</tr>
<tr>
<td>Greet and introduce oneself</td>
<td>“Hi!”</td>
</tr>
<tr>
<td>Express positive emotions and emoticons</td>
<td>“yay! ☺”</td>
</tr>
<tr>
<td>Appreciate team member</td>
<td>“Good job Peter”</td>
</tr>
<tr>
<td>Team Commitment (TCM)</td>
<td></td>
</tr>
<tr>
<td>Express confidence in own team’s ability</td>
<td>“We have the longest chat”</td>
</tr>
<tr>
<td>Show togetherness through ‘We’ language</td>
<td>“Let us do it”</td>
</tr>
<tr>
<td>Hold own team in higher regard</td>
<td>“We are better”</td>
</tr>
</tbody>
</table>

Table A2. Rules defined to isolate nc lines

<table>
<thead>
<tr>
<th>Spam type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spamming with emoticons, characters and sentences</td>
<td>:dislike::dislike:;”</td>
</tr>
<tr>
<td></td>
<td>“ttttttttttttttttttt”,</td>
</tr>
<tr>
<td>Spamming with random characters or other languages</td>
<td>“sdf jkl jk”,</td>
</tr>
<tr>
<td></td>
<td>“ł Ł Ą ĄŁ Ą Ł”,</td>
</tr>
<tr>
<td></td>
<td>“jadi sekarang melakuka”</td>
</tr>
<tr>
<td>Talking about other teams or their physical locations</td>
<td>“im in the other room”,</td>
</tr>
<tr>
<td></td>
<td>“John’s group is talking about pokemon”</td>
</tr>
<tr>
<td>Discussion about random stuff out of task or classroom matters</td>
<td>“who likes gangnam style”</td>
</tr>
<tr>
<td>Talking to the admin or technical issues</td>
<td>“Chat Admin tell us who are u???”,</td>
</tr>
<tr>
<td></td>
<td>“its very slow and lagging”</td>
</tr>
</tbody>
</table>

Table A3. List of punctuations tagged

<table>
<thead>
<tr>
<th>Punctuations</th>
<th>Tags</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive emoticons (✿, :like:, :D)</td>
<td>[{pos_emo}]</td>
<td>Shows positive support to team member</td>
</tr>
<tr>
<td>Neutral and negative emoticons (⊙, :o)</td>
<td>[{neut_emo}]</td>
<td>Do not contribute to teamwork dimensions</td>
</tr>
<tr>
<td>@</td>
<td>[{ref_mark}]</td>
<td>Used to tag or refer to a person in a chat in the context of mutual performance monitoring.</td>
</tr>
<tr>
<td>?</td>
<td>[{question_mark}]</td>
<td>Helps to identify if a chat line is a question or an agreement</td>
</tr>
<tr>
<td>*</td>
<td>[{asterisk_mark}]</td>
<td>Used by students at times to show typo correction which indicates improvement of the answer.</td>
</tr>
<tr>
<td>:</td>
<td>[{colon_mark}]</td>
<td>Used by students to consolidate and present the answer to the problem showing co-ordination.</td>
</tr>
<tr>
<td>Hyperlinks http://</td>
<td>[{weblink}]</td>
<td>Instances of providing external references related to the task like websites found in web search</td>
</tr>
</tbody>
</table>
A Multivocal Approach in the Analysis of Online Dialogue in the Language-focused Classroom in Higher Education

Giulia Messina Dahlberg
School of Health and Learning, University of Skövde, Sweden // giulia.messina.dahlberg@his.se

ABSTRACT
The study presented here is interested in understanding the ways in which social interaction in technology-mediated institutional settings is constrained and afforded by what Pennycook defines as “critical moments” in the educational experience. Drawing on Social Learning Analytics and on the concepts of heteroglossia, contingency and chaining, this paper critically discusses a methodology that allows the analysts, and ultimately learners and educators, to follow and visually represent the mobility of the learners-in-concert-with-tools across space, time and language varieties and modalities in technology-mediated communication. The empirical data focused on here is drawn from a large project which includes 40 hours of naturally occurring interactional materials, generated through screen recordings of online sessions of an Italian for Beginners course offered by a Swedish university in the videoconferencing platform Adobe Connect. Preliminary findings suggest that the environment, both in terms of what happens inside and outside the virtual learning site, is of primary importance when it comes to the organization of the interaction among individuals in terms of what becomes the participants’ focus during the encounters. A visual representation through a multivocal approach of this mobility of learners, topics and tools will, it is suggested, support learners, educators and designers in locating where and when in the interaction “critical moments” have occurred, in order to understand how such shifts in focus support or hinder the learning experience.

Keywords
Technology-mediated communication, Videoconferencing, Virtual learning sites, Social learning analytics, Critical moments, Dialogue, (Im)mobilities, Ethnography, Multivocal approach

Introduction
How do we capture those critical moments where something changes, where someone “gets it,” where someone throws out a comment that shifts the discourse? (Pennycook, 2012, p. 132)

As technology-mediated communication (TMC) is increasingly becoming a part of our everyday lives (at least in the global North), we are endowed with the possibility of engaging in communication everywhere we go and with whoever we wish to, without worrying about logistical issues. While TMC is a gloss for communication in and through different kinds of texts, e.g., email and computer applications for videoconferencing, it has interesting implications when issues of how meaning is negotiated in online spaces are addressed. Being inside the virtual classroom and engaging in TMC to interact within a learning community implies that participants need to adjust to the media and artifacts that are dimensions of that space: “It is at the intersection between people’s actions, the tools they use and the infrastructures that they have access to that new conditions for learning arise and where new practices emerge” (Bliss & Säljö, 1999, p. 7).

The concept of social learning analytics as developed by Buckingham Shum and Ferguson (2012) is a subset of learning analytics that focuses on the study of learning in the doing, i.e., group processes and the co-construction of knowledge. According to Buckingham Shum and Ferguson (2012), social learning analytics “should render learning processes visible and actionable at different scales: from national and international networks to small groups and individual learners” (p. 5). In addition, referring to the use of inscriptions and other kinds of technologies as tools that mediate thinking, Säljö (1999) argues that “[t]he mastery of mediational means is […] an essential aspect in the process of learning” (p. 152) wherein a fundamental assumption is that “learning is always learning to do something with cultural tools” (p. 147).

Moreover, Hampel and Hauck (2006) highlight that affordances for participants are related to aspects of the use of “the most appropriate tools” among those offered in digital spaces and those that are best suited to the communication situation participants have at hand. Wertsch (1998) refers to a tension in the sociocultural continuum between agents and their mediational means because mastering the use of tools can both enable and constrain action. Mapping these tensions using the epistemological lenses of social learning analytics is a central interest in this study. In addition, given the ontological assumption in the present project of considering mediated action (i.e., agents and their cultural tools) as the fundamental unit of analysis, how can this element be conceptualized so that it makes sense and can be used to support learning analytics? I argue that a focus on dialogue and its sequentiality in space and time is crucial for understanding the organization of interaction in
Heteroglossia in online dialogue

Bakhtin’s notion of heteroglossia builds upon an understanding of language (and of the world) as social practice in diverse contexts, as opposed to the idea of language as a generic system of linguistic symbols. Dialogue and heteroglossia, in Bakhtin’s terminology, are two aspects of the same phenomenon that cannot be separated since they generate each other in the situated interaction and meaning making through any utterance. Following this line of thought, in dialogue, and also in online dialogue, it is relevant to analyse utterances as an activity of orchestration of the agent who produces the utterance and the mediational means that afford and/or constrain the communicative project (Linell, 2009), and thus the co-production of shared and mutual knowledge. In a similar vein, Shegloff (1996) describes an utterance as interdependent of its interactional past and possible future, in the form of the next utterances. Scholars in the field of conversation analysis (CA) aim to understand discourse in the local sequentiality of situated interaction in terms of utterances’ retrospective dependence (i.e., Sacks, Schegloff & Jefferson, 1974), their focus on the present and their projections towards possible future(s). Such a view is in line with heteroglossic understandings on language since it entails a conceptualization of discourse as an activity accomplished in situ, drawing on the resources that are at hand and focusing on the interlocking links that are constitutive parts in the interaction. Thus heteroglossia, in Bakhtinian terms, is an epistemology of complexity that understands the organization of utterances as a sequential endeavour in time and space (Bakhtin, 1981).

We have in previous work (Messina Dahlberg & Bagga-Gupta, 2013a; Messina Dahlberg & Bagga-Gupta, 2013; Messina Dahlberg & Bagga-Gupta, 2014 Messina Dahlberg, 2015) attempted to provide an account of technology-mediated interaction by means of an augmented CA that takes into account the mediational means afforded by the videoconferencing platform used in some institutional encounters. Using the analytical tools of CA that allow a fine-grained understanding of turn-taking, we were able to map parts of the fractured ecologies (Luff et al., 2003) of technology-mediated communication. Patterns in the data show the range of ways in which interaction in the online environment is affected by the task at hand, as well as how the tools support (or constrain) participants’ communicative projects (Luff et al., 2003; Messina Dahlberg & Bagga-Gupta, 2016). The study presented here is an attempt to develop our previous studies by merging the epistemological stances of social learning analytics, heteroglossia and dialogism. I will present some illustrative examples that shed light on the dynamic and fluid nature of communicative projects that can change as “they are carried out or brought to completion, sometimes in ways that were not projected from the beginning” (Linell, 2009, p. 178). Before presenting the data in such examples, I will, in the next section, provide some background of the study of online dialogue in relation to learning analytics.

Analytics in online dialogue: Some recent studies

In educational contexts, the uncertainty about which direction the dialogue is going to take has been conceptualized in terms of creativity (Wegerif et al., 2010). The alternative pathways of interaction are understood as widening the space of dialogue in critical thinking. Visual representations of the sequentiality of creative dialogues in Wegerif’s study are created by a computational model that recognizes dialogic creativity in terms of “new perspectives.” The study’s sequence diagrams that represent online discussions in a graphical text-based online environment called Digalo (see also Schwartz, Dreyfus & Hershkowitz, 2009) offer a powerful visualization of the instances where new perspectives emerge in the discussion. Results of the analysis of the explorations of the computational model highlight how the latter can offer relevant and reliable insights into the
ways in which meaning is created in dialogue by participants. Furthermore, the spatial representation afforded in Digaloo maps allows “pedagogical affordance for creativity” (Wegerif et al., 2010, p. 620). Scholars have reported on the design and use of online tools that build upon slightly different ontological assumptions of “good dialogue” and deliberations, compared to the focus on creativity in the study by Wegerif et al. (2010). Here, focus lies on exploratory dialogue and the assessment of the quality and quantity of the evidence that support a given participant’s argumentation (De Liddo & Buckingham Shum, 2013) as well as on participants’ epistemic beliefs when dealing with collaborative tasks (Knight, Buckingham Shum & Littleton, 2013a). The concepts of epistemology, assessment and pedagogy and their relation to issues of analysis and representation of human interaction, epistemic practices and higher-order thinking with learning analytics are discussed by Knight et al. (Knight, Buckingham Shum, & Littleton, 2013a; Knight, Buckingham Shum, & Littleton, 2013b; Knight & Littleton, 2015). From the epistemological assumption that “our learning analytics are our pedagogy” (Knight Buckingham Shum & Littleton, 2014, p. 31), the authors bring to the fore the burning issue of the purpose of designing and using learning analytics in educational research. Taking pragmatic and sociocultural perspectives, Knight, Buckingham Shum and Littleton (2014) highlight the importance of discourse-centric learning analytics in order to capture the role of the performativity of epistemic practices, i.e., the knowledge-in-the-making by participants across contexts and technologies-in-use. According to Knight and colleagues (Knight, Buckingham Shum & Littleton, 2014; Knight & Littleton, 2015) learning analytics is a tool that frames a certain kind of analysis and not a goal per se that shapes what the idea of meaningful learning should be.

According to Mercer and colleagues (Mercer & Wegerif, 1999; Littleton & Mercer, 2013), collaboration and co-reasoning in educational contexts are also framed in terms of exploratory dialogue in Ferguson and Buckingham Shum (2011). In an attempt to identify exploratory dialogue as well as resources that support learning within synchronous textual chat, Ferguson and Buckingham Shum provide a preliminary analysis of the contributions in the chat tool included in the web conferencing platform called Elluminate. The exploratory markers identified in the transcripts of the discussions indicate the importance of considering the context of the contributions, since there seems to be a correlation the position of the messages in time and the amount and the quality of meaningful exchanges among participants. In their later evaluation of the study, Ferguson et al. (2013) propose “a self-training framework for the detection of exploratory dialogue within online discussion” (p. 92) as well as a number of visual applications of computational models for exploratory dialogue detection. Also in this later study, Ferguson et al. (2013) re-confirm the importance of analyzing the context of the utterances in order to be able to identify instances of meaningful exchanges (within the epistemological frame of exploratory dialogue) in their sequentiality in time and space. Suthers et al. (2010; 2013) propose an analytical framework that accounts for the co-occurrences (or, as it has been framed elsewhere, the “chaining” [Bagga-Gupta, 2002; Bagga-Gupta, 2015; Messina Dahlberg & Bagga-Gupta, 2013]) between what they frame in terms of events, activity, uptake and contingency in the study of human interaction (see Figure 1).

Contingencies are framed in terms of “how acts are manifestly related to each other and their environment” (Suthers et al., 2010). Such a focus on the importance of context in terms of the sequentiality of the utterances and their relation to their environment is in line with the deliberations in the Introduction section of this paper where the concepts of heteroglossia in dialogue and the methodological assumptions of CA were outlined. In the following sections, I will provide an example of how learning analytics could be used to visualize the (im)mobilities of participants’ embodiment, tools and technologies-in-use in videoconferencing, more specifically when students are engaged in using (and learning) a language variety with which they have limited experience as well as the use of a range of communicative tools (both digital and analogue, e.g., written chat, whiteboard, course materials and hand-written notes) both inside and outside the virtual environment of the videoconferencing platform Adobe Connect. Special focus lies on the chaining, or contingencies of people-in-concert-with-tools as well as on the attempt to discern critical moments in the interaction where participants
change perspective and focus. Such a focus on context(s) and boundaries will, it is suggested, provide an alternative voice in the fields of discourse-centric and social learning analytics.

**An example: (im)mobilities in online dialogue**

In the previous sections, I have briefly reviewed the Bakhtinian concept of heteroglossia as well as the notion of chaining and how these frame the understanding of dialogue and situated cognition and interaction in this paper. One important issue at this stage is how the epistemological assumptions of such a sociocultural-dialogical take on interaction are mirrored in the creation of the social learning analytics that will eventually shape our pedagogy in terms of what epistemic practices or “learning experiences” could be.

This paper draws upon the ongoing work in the CINLE project (Everyday Communication and Identity Processes in Netbased Learning Environments, see http://ju.se/cd/cinle) in which two Italian for Beginners online courses are examined. Recordings of the synchronous online meetings (once a week during two terms, for a total of 40 hours of interactional material) as well as the course instructors’ planning and materials are included in the data (see also Messina Dahlberg [2015] for a detailed description of the data presented in this paper with illustrative purposes). CINLE aims to develop knowledge regarding the situated practice of interaction in institutional learning settings where online language courses are offered at university level. Taking sociocultural and postcolonial perspectives, we illustrate the ways in which learning, culture and identity are framed within the affordances and the constraints of the online environment, as well as how an epistemology of time/space as a single dimension is fruitful for understanding: (i) how the synchronous sessions are organized based on the analysis of the interaction; (ii) how tasks and activities are chained in the situated interaction; (iii) how literacy practices frame the interactional organization; and, finally, (iv) to highlight how the (im)mobilities of tools and participants have a bearing on the time/space organization in the online environment. (Im)mobility here refers to the prerogative of online courses in which students may be required to adhere to a specific schedule and participate online at specific temporal slots, but there is no need for them to congregate at a given physical place.

**CA representation of online dialogue**

The virtual classroom in which data for the CINLE project has been generated consists of a videoconferencing program that allows participants to use oral and written communication as well as share their individual webcams. A critical aspect of TMC in general lies in its very mediational component: by means of the digitalization of the processes at stake in the online interaction, the tools and the inscriptions inside the virtual classroom become visible and can be accounted for. The analyst can keep track of the documents displayed on the whiteboard and the interaction that occurs in the written mode in the chat pod during the meeting, as well as record the participants’ contributions in the oral mode. The researcher has the vantage point of uncurtailed access to everything that occurs inside the environment and can thus, a posteriori, access the contributions in the environment in all the modes to map how these are mutually shaped in interaction.

In the following sections, I will provide some illustrative examples of what this access to a range of epistemic practices means in terms of representation and analysis in order to create learning analytics that account for such a movement across modalities, language varieties and a range of other mediational means.

**Tracing chaining in online dialogue**

*Zooming in: Interactional sequence and adjacency-pairs as units of analysis*

In order to illustrate these issues, one instance is examined which took place during one synchronous session of the online course Italian for Beginners where participants’ contributions shape and are shaped by the inscriptions inside and outside the environment of videoconferencing. It is important to highlight that in order to make visible participants’ use of different language varieties, all original language use has been retained in the transcriptions and a verbatim translation in English is provided (see Figure 7).

In line 03 (of Figure 2), Olle refers to the file with the questions about the topic of the day, (see Appendix). In line 05 Olle reads the questions, after specifying in Swedish: *de där frågorna står ju då* (Original in Swedish [Sw]: these are the questions then). Olle takes on the task of the “reader of the questions” that are at this point, after the move to the small group space, not available on the WB in the virtual environment. Olle is accorded this
position by Anna, who in line 01 asks him whether he wants to start. After a silence of 5 seconds, Anna adds: *om du vill* (Sw: if you want). Here silence appears to be a consequence of the medium where the session is taking place. Participants cannot be certain that the other members have understood their previous turn-at-talk. Here one can consider understanding or epistemic moves as embodied achievements in interaction (Mondada, 2011). This shows that participants have access to interlocutors’ bodies through oral markers like the participants’ voices and the lexical contents of the contributions. These processes are constrained by the limited visual access that the participants have to one another in the online environment, thus resulting in long moments of silence and in the disruption of the indexical order of the conversation. Olle orients towards this marker of uncertainty displayed by Anna in line 01 when he leaves the floor to her in line 10.

Figure 2. Student’s task orientation

Figure 3 below presents the interactional order that can be seen in student-only phases of online contexts. Here the interactional sequence continues on the topic of friends. Olle has previously described his friend in the task-oriented activity. Olle’s friend has, according to Anna, a Dutch-sounding name (Anna is physically located in the geopolitical spaces of the Netherlands during the meeting). The sequence begins with a question, where Anna asks whether Olle and his friend speak Dutch together.

Olle’s hesitation with regard to Anna’s question is highlighted (in Figure 3) by: (i) the silence in line 42; (ii) the *e:m* produced by Olle in line 43; and (iii) the overlapping talk in 44 where Anna adds *o inglese?* (Original in Italian [It]: or English?). Olle takes the turn in line 45, thus initiating the epistemic engine of the sequence that will end in line 59: *qualche volta parliamo olandese* (It: sometimes we speak Dutch). Olle looks for the word item sometimes in Italian, positioning himself in a lower epistemic status. Anna orients towards Olle’s turn and produces an answer with rising intonation as a marker of her uncertainty, followed by a repair in line 50. Olle continues the word search in line 51 using declarative syntax. The analysis highlights a general pattern in the data, where students use other language varieties (in this case, Spanish) when they are uncertain about a particular word item which does not allow them to carry on the conversation. This transpires particularly when producing the correspondent word item(s) in Swedish does not help in the search either (see line 46, where Olle produces the Swedish word *ibland* [Sw: sometimes]). Per, the third participant in the meeting, utters some word items rather loudly in the overlapping talk depicted in line 53. Per repeats *qualche volta* (It: sometimes) a second time in line 58 after Olle’s request for clarification in line 56. After a pause, Olle is finally able to produce the sentence he initiated in line 45. In concomitance with line 64, the same sentence is made publically available by Per in the chat pod of the environment. This is also a rather common behaviour that has been identified in the
data: students use the chat window either to perform parallel conversations or as an alternative to the WB in the environment, where they can write word items or other information related to what has been raised in the oral modality or vice-versa. The analysis shows further how the turns in this sequence are finely tuned to one another: Anna orient towards Olle’s hesitant talk in line 62 by filling in the word item la lingua (It: the language) in the overlapping talk illustrated in line 63. Olle uses the same word item in the latching turn in line 64. Anna’s repetition, in lieu of confirmation in line 67, concludes the sequence, after which Anna orient to Per and asks him about his friends. The repetitions, both in the oral and in the written modes, of word items produced by the participants during the online meeting are understood as a result of the limited visual access to one another: Per writes the same words he produced twice, the second time very clearly, in the oral modality in order to “put into print” what has been said before and that the other participants were looking for in their word search. Anna and Olle’s mutual fine orientation shown throughout the sequence is, I maintain, also a consequence of this lack of visual access to one another’s bodies or the non-embodied characteristics of this interaction.

<table>
<thead>
<tr>
<th>Time</th>
<th>Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>06:43</td>
<td>Anna: e: do you speak dutch together?</td>
</tr>
<tr>
<td>06:43</td>
<td>Olle: o: [throat clearing-laugh]</td>
</tr>
<tr>
<td>06:43</td>
<td>Anna: o: do you say av-aveses aveses e: vad how do you say av-aveses aveses e: how heter det [bland]! do you say sometimes!</td>
</tr>
<tr>
<td>06:43</td>
<td>Olle: e:</td>
</tr>
<tr>
<td>06:43</td>
<td>Anna: e:</td>
</tr>
<tr>
<td>06:43</td>
<td>Olle: e: nej inte kanske [bland] aveses heter det no not maybe [sometimes] aveses &gt;it’s called</td>
</tr>
<tr>
<td>06:43</td>
<td>Per: =QUALCHE VOLTE SOME TIMES</td>
</tr>
<tr>
<td>06:43</td>
<td>Olle: =a-a-a</td>
</tr>
<tr>
<td>06:43</td>
<td>Anna: qualche volta[ ] some time</td>
</tr>
<tr>
<td>06:43</td>
<td>Olle: =ad heter det sa du? =w:hat is it called did you say?</td>
</tr>
<tr>
<td>06:43</td>
<td>Per: =QUALCHE VOLTE SOME TIMES</td>
</tr>
<tr>
<td>06:43</td>
<td>Olle: ah qualche-qualsche volto: [parliamo: ] olandese ah some-some times we speak dutch</td>
</tr>
<tr>
<td>06:43</td>
<td>Anna: =ah olandese he: maxe: max: [e: e: e: e: e:] spre = the dutch language he: but but but of</td>
</tr>
<tr>
<td>06:43</td>
<td>Olle: =spesso parliamo: inglase</td>
</tr>
<tr>
<td>06:43</td>
<td>Anna: e: often we speak english</td>
</tr>
<tr>
<td>06:43</td>
<td>Olle: inglase</td>
</tr>
<tr>
<td>06:43</td>
<td>Anna: english</td>
</tr>
</tbody>
</table>

Figure 3. Chaining across modalities
At a more overarching level, an analytical gaze at Figure 2 also highlights how Olle’s orientation to the online and offline resources during his on-task activity implies a lack of orientation towards other resources in the environment, including other participants. This pattern of rather long chunks of talk is disrupted when participants asked direct questions of one another (Figure 3). Figure 3, and the arrows included in it, show the epistemic ‘seesaw motion’ (Schwartz et al., 2009) initiated by participants’ focused mutual orientation in the oral mode. Such movement and mutual orientation is what I have framed in section 1 in terms of (online) dialogue and heteroglossia. In the next section, I will attempt to show this shift in focus by zooming out and focusing on an analytical macro-scale of the interaction.

Zooming out: Before and after critical moments

The ethnographic study of the online interaction across time allows for a preliminary analysis on a macro-scale to attend to interactional patterns that occurred across sessions over one term of study. As illustrated in the analysis of instances from one session during the online course in the previous section, participants clearly initiate their contributions by using vocatives: Olle, would you like to start and if you would like to start with that Anna (Figure 2, lines 01 and 10). These discourse markers are used to clearly define the start and the end of each contribution. This communicative strategy enables a smooth transition and the next student can take the floor (this structuring device enables participants to get turns-at-talk). Shiffrin argues that discourse markers gain their pragmatic function in relation to “the underlying cognitive, expressive, textual and social organization of discourse” (Shiffrin, 2003, p. 66). During the teacher-led online meetings, the students construct a time-space dimension for their contributions by clearly marking the boundaries of the start and end of each turn. Thus, the communication that occurs in the environment can, in terms of the organization of the interaction, be understood as individual presentations by the students who are usually well prepared, rather than as a discussion or a conversation. There seems to be a specific interactional order, a “culture” inside the group: when one student holds the floor, the others are silent and they seldom ask follow-up questions. What they do, thereby preserving the indexical order of the interaction, is that they sometimes relate to previous students’ presentations, by saying, come ha detto Anna… (It: As Anna said…).

The students’ contributions in the oral mode are represented in Figure 4 as chunks (boxes) of similar length that follow one another, with rare overlaps. Red arrows indicate chaining of the contributions across modes and literacy events while blue arrows indicate mutual orientation among students’ turns. Figure 5, on the other hand, visualizes the talk-in-interaction when students engage in dialogue after a critical moment as highlighted in the previous section of this paper. Here, the contributions from the participants are an integral part of the overall oral conversation. After a critical moment, participants respond and react to one another’s contributions with overlapping talk. They commonly ask for clarifications and confirm their mutual attention (see also Figure 3). The length of the chunks of talk thus varies according to the response students receive from one another.
on the ethnographic understanding of the data in CINLE) aims to provide a first step in the implementation of learning analytics that attends to the length of talk, silence, students’ mutual orientation as well as to the tools they have at hand, inside and outside the online environment. Figures 4-5 show the patterns of contingencies (Suthers et al., 2010) or chaining across online/offline resources as well as across participants’ contributions in the oral mode. However, the jump from data to abstract representations (Figures 4-5) using learning analytics must go through a detailed analysis at the micro-scale (Figures 2-3), and this is the analytical challenge that this study addresses, in terms of a methodology that is able to handle such analytical scale-jumps, i.e., the movement from the general to the specific, or from the collective to the individual in the study of human interaction.

Heteroglossia and chaining in discourse-centric social learning analytics

The illustrative examples taken from one synchronous session in an online language course provided the opportunity to discern some of the interactional patterns that are at stake when people interact in institutional settings like the one focused on in this study. Furthermore, the representation on a macro-scale of the analysis of such interactional patterns (Figures 4-5) before and after the occurrence of critical moments illustrates participants’ orientation towards one another as well as across tools and modalities. This illustrative example will, it is suggested, provide the opportunity to create an automated tool for computational analysis that, on the basis of learning analytics, can support a deeper understanding of the (im)mobilities of learners-in-interaction-with-tools in the language-focused virtual classroom. This will allow for a bridge between several analytical scales in the study of synchronous TMC as well as of the dynamics behind the occurrences of critical moments and their educational implications for language learning and instruction.

Figure 6. Multivoval approach in the analysis of online dialogue in the language-focused classroom

Figure 6 is an attempt at detecting the relevant analytical scales that provide a depth in the analysis to support a visual representation of online dialogue. A multivoval analysis (Suthers & Rosen, 2011) takes into account and shows the chaining across scales, modalities and language varieties as well as the tools used by participants in the online/offline settings. More specifically, I contend that an analysis at the micro scale would inform the multivoval endeavour of following, mapping and measuring illustrated in Figure 6 by focusing on:

- adjacency pairs or responses (see Goffman (1981) and Figure 3)
- chaining across modalities through repetition or ventriloquizing (see Tannen (2007) and Figures 2, 4-5)
- length of turns-at-talk, overlapping talk and pauses (see Figures 4-5)
- occurrence of novel word-items/chunks in the target language (see Messina Dahlberg and Bagga-Gupta [2013; 2014] and Figures 2-3)
- epistemic (im)balance across participants’ enquiring initiations (K-) and knowing utterances (K+) (see Heritage (2012) and Messina Dahlberg & Bagga-Gupta (2016))
- discourse markers, e.g., start and end of utterances (see Shiffrin (2003) and Figure 3)

Such a take on analytics forefronts epistemologies that attend to the learning potential of heteroglossia in online dialogue, e.g., a scientific position that acknowledges the complexity of discursive practices in (online) interactional (learning) spaces. The analysis at the micro-scale shows how the unexpected turns (or critical moments) that emerge in interaction can be understood as driving the dialogue forward and are therefore of primary interest for mapping and studying using analytics. In addition, a further dimension of language use
needs to be addressed here, i.e., participants’ manipulation of word items and chunks related to tasks in the course materials as well as in the course curriculum more broadly. The implementation of a multivocal approach to the creation of a more holistic analysis of online (synchronous) group interaction means providing an interpretation of the data which is in dialogue with other interpretations. This is done in order to reach the goal of the creation of alternative theoretical and methodological understandings of the field as well as the data which can then be generalizable and usable as a method for reliably distinguishing “critical moments” from non-critical moments, not least with regard to issues of causality that are often problematic in educational research, not least concerning learning analytics which rely precisely on a relation of cause-effect: “The purpose of a scientific learning analytics is to identify underlying causal mechanisms and provide actionable advice that pertains directly to learning.” (Rogers, 2015, p. 228). I share Rogers’ (2015) concerns about the endorsing of mixed methodologies not “as an act of methodological tolerance [but rather] from a point of ontological validity” (p. 229). Such concerns and discussions, from a critical realism perspective in Rogers (2015), or the post-humanist perspective in Barad’s (2003) agential realism are crucial in order to understand the epistemological grounding on which learning analytics are resting. Both concepts entail a shift that is tectonic in nature: the focus no longer lies on science as something that is, and that can be observed, but rather on the means and measurement tools that are used to make science possible.

Conclusions

This paper critically discussed a methodology that allows the analysts, and ultimately learners and educators, to follow and visually represent participants’ dialogue-in-concert-with-tools across space, time and language varieties in TMC. Using the analytical lenses of heteroglossia and chaining, the complexity of online dialogue and the chaining across modalities/tools and language varieties were addressed. Social learning analytics and, in particular, discourse-centric learning analytics, are understood as tools of analysis that could address such complexity. Nevertheless, the illustrative examples of online dialogue in a synchronous videoconferencing setting show that an analysis at different scales is needed that uses a range of methodological and theoretical understandings of what learning, and thus learning analytics, is. I started by taking the Bakhtinian concept of heteroglossia and discourse and its sequentiality as a starting point and as a unit of analysis. Here, dialogue is centre-staged in order to shape a type of learning analytics that will capture moments of shift in focus, or critical moments, after which there is a change in the overall interactional pattern, as I have outlined in section 1 in this paper and which, I maintain, retain specific educational interest in terms of the language use that emerges in concomitance with such shifts (see also Messina Dahlberg & Bagga-Gupta, 2016, for a detailed analysis of these phenomena). The purpose of designing such an analytical tool and defining a more fluid unit of analysis that depends on the power of magnification used in looking at the data and the issues that are at stake, would be to create a space for dialogue across disciplines and an alternative to other discourse-centric learning analytics. The latter, it is suggested, takes dialogue as an a priori defined standard of productive talk, in terms of the horizontal (higher/lower level thinking) and vertical (creativity) movements as outlined in the studies reviewed in the section called Analytics in online dialogue: some recent studies. Access to large amount of information (in terms of so-called “big data”) is the analytical centre in the fields of educational data mining and learning analytics. Here, the digital traces that students leave behind are used in the analysis in order to study “learning in the doing.” According to Buckingham Shum and Ferguson (2012), these kinds of analyses “should render learning processes visible and actionable at different scales: from national and international networks to small groups and individual learners” (p. 5). Analytics thus, always entail some kind of preparation of data for further analysis. This, in turn, means that issues of standards, quality and globalization are at stake now more than ever. I argue that, in order to address this issue, empirically grounded research at the micro scale of analysis is needed in order to understand the ways in which interaction may support collaboration and learning across time and space (see also Chen et al., 2016). As Suthers and Rosen (2011) succinctly put it: “the network structure is not enough: to explain the origin of social life we must understand the nature of the communication or interaction that takes place” (p. 17). The organization of time and space in digitally mediated interaction in specific virtual learning sites both affords and constrains the emergence of critical moments in which participants experience a shift in focus that, in turn, may end up in an encounter with new knowledge and a reorganization of the flow of the interaction. The methodological analytical deliberations in this study will, I argue, contribute to the field of language learning and instruction, where learning analytics can be designed to attend to issues of manipulation of novel structures and word items in the target language in the language-focused virtual classroom.

The implementation of a multivocal approach to the creation of a more holistic analysis of online group interaction as outlined in the study by Balacheff and Lund (2013) means providing an interpretation of the data which is in dialogue with other interpretations, not necessarily to create a common understanding (or convergence in Balacheff and Lund’s terminology, see also Rogers (2015)) but with the aim of developing
“further theoretical and methodological integrations” or to “progress in each other’s problématiques” (Balacheff & Lund, 2013, p. 11). However, the data that has been used with illustrative purpose in the present study has been generated through ethnographic methods. This implies a more holistic understanding of the field, which is a vantage point for the researcher(s) who conducted the work. Such a prerogative needs to be shared in appropriate ways in dialogue with other researchers in order to reach the goal of the creation of alternative theoretical and methodological understandings of the field as well as the data. The co-creation of mergers, coalitions and challenges in the analysis of the data through a multivocal approach invokes the establishment of some sort of shared focus, as advocated by Balacheff and Lund (2013), in which data is seen as the boundary object between learning analytics and educational data mining as different approaches to online learning. In the present study, the boundary object, seen as an artifact that articulates meaning and addresses different perspectives (Akkerman & Bakker, 2011) is participants’ learning space(s), included in the online videoconferencing platform. Such spaces include artifacts that emerge from a social network which is in constant flux or a process of (re)making, wherein participants are continuously orienting towards a range of tools and practices, enabling a “move” between different virtual and real (or online and offline) spaces and the management of boundaries. Since such a movement can be conceptualized in terms of “implicit transitions,” boundary objects are, at least for the individuals who are dealing with a range of tasks and activities in it, black-boxes. People thus engage with them from a taken-for-granted position. A multivocal approach to the analysis of online dialogue aims at opening up such black-boxes. This is a high-stakes endeavour, which seeks to understand not what the ingredients of productive interaction are, but rather, to unfold the educational practices with which participants get engaged or disengaged, or, as Säljö (2012) puts it, “as an alternative to resting content […] in terms of abstract outcome measures of products of learning” (p. 12).

### Transcription key

| Underlined | denotes that the word is focally accentuated |
| ? | rising intonation, not necessarily a question |
| ! | strong emphasis, with falling intonation |
| () | micro-pause |
| (1) | pause in seconds |
| (laugh) | verbal description of actions noted in the transcript, including non-verbal actions |
| (xxx) | indicates a stretch of talk that is unintelligible to the analyst |
| e: | one or more colons indicate lengthening of the preceding sound |
| :m | a hyphen indicates an abrupt cut-off, with level pitch |
| SOME | indicates loud voice |
| = | indicates that there is no gap at all between the two turns |

**In the English translation**

- *Italics* original utterance in Italian
- *Bold* original utterance in English
- *Bold* original utterance in Spanish

- Indicates chaining between turns in the same mode
- Indicates chaining between turns in the different modes (from oral to written)
- Indicates chaining between turns in the different modes (from written to oral)
- Indicates chaining between oral contributions and WB
- Indicates chaining between oral contributions and one student’s offline notes
- Indicates chaining between oral contributions and course material displayed on the WB
- Indicates chaining between oral contributions and course material not displayed on the WB

**Figure 7. Transcription key**
References


## Appendix

### Questions in preparation of the meeting in Figure 1 and 2

<table>
<thead>
<tr>
<th>Före chatten kommer du att förbereda kapitel 14 i boken Buon viaggio Vi kommer att prata om VÄNNER</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Presenta un tuo amico: come si chiama, quanti anni ha, che cosa fa…</td>
</tr>
<tr>
<td>• Parla del suo carattere: come è?</td>
</tr>
<tr>
<td>• Incontri spesso i tuoi amici?</td>
</tr>
<tr>
<td>• Cosa fai insieme a loro?</td>
</tr>
<tr>
<td>• Tu e i tuoi amici avete gli stessi interessi? Quali?</td>
</tr>
<tr>
<td>• Preferisci stare solo/sola con una persona o con una grande compagnia di amici?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Before the meeting you shall prepare chapter 14 in the book Buon viaggio. We are going to talk about FRIENDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Introduce a friend of yours: what is his/her name, how old is he/she, what does he/she do…</td>
</tr>
<tr>
<td>• Talk about his/her character: how is he/she?</td>
</tr>
<tr>
<td>• Do you often meet with your friends?</td>
</tr>
<tr>
<td>• What do you do with them?</td>
</tr>
<tr>
<td>• Do you and you friends have the same interests? What?</td>
</tr>
<tr>
<td>• Do you prefer to be alone with a person or with a big group of friends?</td>
</tr>
</tbody>
</table>

The Effects of Cognitive Styles on the Use of Hints in Academic English: A Learning Analytics Approach

Sherry Y. Chen* and Chia-Chi Yeh
Graduate Institute of Network Learning Technology, National Central University, Jhongli, Taiwan // sherry@cl.ncu.edu.tw // comkingl@yahoo.com.tw

*Corresponding author

ABSTRACT
Various hints have been proposed to scaffold students. Such hints can be broadly divided into direct hints and indirect hints. On the other hand, individual differences existed among learners. In particular, cognitive styles greatly affect student learning. However, there is a lack of studies to investigate how cognitive styles affect students’ reactions to the use of hints in the context of academic English. To address this issue, this study aimed to examine the effects of cognitive styles on the use of hints in such a context. To obtain deep understandings, a learning-analytics approach was applied in this study, including quantitative measurement, qualitative evaluation and lag sequential analyses. The results from the quantitative measurement indicated that the frequencies of using Chinese hints and synonym hints were significantly associated with Serialists’ task scores while such significant correlations were not found for Holists. The results from the qualitative evaluation showed that Holists favored to use synonym hints while Serialists preferred to use Chinese hints. The results from the lag sequential analyses suggested that Holists demonstrated iterative behavior while Serialists showed sequential behavior. In summary, cognitive styles have considerable influences on students’ learning patterns in the context of Academic English.

Keywords
Cognitive styles, English learning, Hints, Lag Sequential Analysis

Introduction

Theoretical background

English has been an international language so Asia schools consider English as the first “foreign” language (Yang & Zhang, 2015). Thus, how to enhance English abilities is a critical issue in non-English speaking countries because such abilities affect the overall competition ability of a country (Chen, Hsu, Li, & Peng, 2006). However, previous research found that English learning may be a challenge to students (Ombati, Omari, Ogendo, Ondima & Otieno, 2013). On the one hand, a universal problem that language teachers raise is that most learners feel reluctant to read English materials on their own (Arnold, 2009). On the other hand, some researchers attributed the reasons for such a problem to the lack of learners’ motivation and self-efficacy (Milková & Hercik, 2014). Moreover, Wu, Huang, Chao, and Park (2014) indicated that learners may not be able to construct relationships between different vocabularies and may not possess abilities to deduce, analyze, gauge, organize, or apply such vocabularies. This may be due to the fact that learners usually learn new vocabularies by rote memorization, which often causes boredom (Min & Hsu, 2008).

A number of approaches can be applied to solve these problems. Among them, E-assessment is a popular activity adopted by teachers (Wang, Li, Feng, & Liu, 2012). E-assessment, which refers to the evaluation of the knowledge or skills of learners in a computer-based environment, has received much attention in educational settings (Fox, 2013). For instance, Hwang and Chang (2011) incorporated E-assessment into a local culture course and examined relationships between learning interests and learning attitude. 61 students, who participated in their study, learned the historical background and the transit of local culture via E-assessment. The results indicated that learners’ learning interests and learning attitude in the culture course were greatly improved after they used E-assessment. Additionally, Zlatović, Balaba and Kermek (2015) employed E-assessment to support an informatics course and investigated students’ learning strategies and learning performance. 351 learners from higher education institutions, who enrolled in an introduction to informatics course, participated in their study. The results showed that the E-assessment not only had positive impacts on their learning strategies, but also improved their learning performance.

In addition to the aforementioned advantages, the E-assessment also has other advantages, including improved question styles using interactive multimedia technology (Linden & Glas, 2010), improved validity and reliability (Xiao & Lucking, 2008) and helpful for learners to review what they have learnt (Butler, Karpicke, & Roediger, 2007). In brief, using E-assessment to support learning is helpful. Nevertheless, E-assessment also has some problems. For instance, it only emphasizes on whether learners can understand and remember what they have learnt out of the context (Sánchez-Vera, Fernández-Breis, Castellanos-Nieves, Frutos-Morales, & Prendes-...
Espinosa, 2012). In other words, such a process does not generate new teaching opportunities for teachers and new learning opportunities for learners. Actually, ideal assessment should be more comprehensible and contribute to make new opportunities for teaching and learning. The other problem is that learners’ anxiety still exists. Such anxiety may negatively affect students’ learning motivation or learning performance. To address these issues, there is a need to provide learners with scaffolding, which can reduce their anxiety and let them work well under pressure (Hsieh, Jhan, & Chen, 2014). Scaffolding refers to support or guidance provided by a mechanism so that learners will not think that a task is too difficult to effectively complete (Belland, 2014). This is because scaffolding uses dialogue to help learners develop ideas they most likely would not have had on their own but such ideas are recognized as the outcome of their own thought (Game & Metcalfe, 2009).

Motivation and aims

Scaffolding can be presented as several ways, among which hints are one of the most useful ways (Devolder, van Braak, & Tondeur, 2012). Hints act as mediators between a student and knowledge he/she attempts to understand and ultimately assist him/her in reaching goals, instead of being accomplished by himself/herself alone (Khaliliaqdam, 2014). Bannert, Sonnenberg, Mengelkamp, and Piegler (2015) claimed that hints can help learners analyze a problem and make a schema to explore a possible solution. Moreover, Bull, Shuler, Overton, Kimball, Boykin, and Griffin (1999) indicated that hints are not one-way, but interactive and reciprocal process between students and the source of instruction. During such a process, students are not passively receiving assistance but actively engaged in the learning process to benefit from the hints to attain a higher level of achievement (Rogoff, 1990).

Because of these potential, various hints have been proposed to support student learning and such hints can be broadly divided into direct hints and indirect hints (Bannert & Reimann, 2012). The former might help students get right answers but they might not be able to develop good reasoning. Conversely, the latter might not be able to help students get right answers but they have potential to help students develop good reasoning. The direct hints are useful to learn several facts while the indirect hints are helpful to understand some concepts. In other words, these two types of hints have different advantages and disadvantages.

On the other hand, learners have diverse knowledge, skills and background so they may favor different types of hints. In other words, there is a need to take into account individual differences. As suggested by Stone (1998), individual differences in cognitive, linguistic and interpersonal skills would affect the effectiveness of scaffolded instruction. Nonetheless, there is a lack of research into relationships between the use of hints and learners’ individual differences (Paas, Tuovinen, Tabbers, & van Gerven, 2003). Among various individual differences, current research found that learners’ cognitive styles influenced how they interacted with technology-based learning (Chen & Macredie, 2010).

Within the area of cognitive styles, differences between Holists and Serialists have been paid attention. According to Jonassen and Grabowski (1993), Holists focus on the content of comprehensive framework while Serialists emphasize on the details of the contents. Ford and Ford (1993) found that Holists and Serialists favored different navigation tools. Holists preferred to use a hierarchical map while Serialists favored to use an index. In addition, Clewley, Chen, and Liu (2011) indicated that Holists and Serialists showed different behavior patterns. Holists tended to use hyperlinks to browse information whereas Serialists favored to use the Previous/Next button to browse information step by step. Subsequently, Chan, Hsieh, and Chen (2014) examined how Holists and Serialists used electronic journals. The results showed that Holists and Serialists used different ways to judge the relevance of documents. More specifically, Holists tended to use a variety of approaches while Serialists preferred to use a single approach. In brief, the results from previous research demonstrated that Holists and Serialists have different learning preferences.

However, such research ignored the reactions of Holists and Serialists to the use of various hints. Thus, this study attempted to fill this gap. In other words, this study aims to examine how Holists and Serialists react differently to the use of various hints. Among a variety of topics, hints are used to support students to learn academic English in this study. This is due to the fact that academic English is essential for the dissemination of research outputs and the establishment of international collaboration. In summary, this study makes contributions to language learning in two aspects. One is to use hints as scaffolding instruction to support students to learn academic English in the context of e-assessment. The other is to contribute to the knowledge of the effects of cognitive styles on the use of hints in an academic English course.
To deeply disclose such contributions, we adopt a learning-analytic approach, which refers to the interpretation of a variety of data created by and collected on behalf of learners to identify potential issues (Long, Siemens, Conole, & Gašević, 2011). Hence, an integrative data analysis was applied in this study. More specifically, both quantitative and qualitative methods are applied to analyze data. Furthermore, we also investigate students’ behavior with a Lag Sequential Analysis, which has been recognized as a promising approach that can identify significant sequential behavior relationships in a series of behavior. Hou, Chang, and Sung (2007) found that the Lag Sequential Analysis can not only identify significant sequential behavior patterns but also illustrate the relationships between such behavior patterns with visual diagrams. Due to such benefits, the Lag Sequential Analysis is also employed in this study. By doing so, we can obtain a complete understanding of (1) how different cognitive style groups perform, (2) what their behavior differences are, (3) why they behave differently, and (4) when they can perform well.

**Methodology**

**Design**

The development of an online test

We developed an online test to help students learn Academic English, which is important to research students because it can help them develop an international perspective (Yang & Zhang, 2015). In particular, the content focused on enhancing students’ understandings of English phrases.

To help the research students learn English phrases without frustration, the design principle of the online test was scaffolding instruction (Wood, Bruner, & Ross, 1976), which means to provide support for students so that they can achieve a deeper understanding for a complex task (Irvin, Meltzer, & Dukes, 2007). Due to such an advantage, scaffolding instruction has been widely used in English learning (Dare & Polias, 2001). Consequently, we also incorporated scaffolding instruction into the online test, where hints were applied to scaffold students. More specifically, the online test provided two types of hints, i.e., synonym hints and Chinese hints. The former, which were indirect hints, provided the meanings of English phrases in English synonyms while the latter, which were direct hints, translated the meanings of the English phrases into Chinese. Based on such hints, the online test had three versions. One version provided multiple hints (i.e., the MH condition, Figure 1), another version contained no hints (i.e., NH condition, Figure 2), and the other version included Chinese hints only (i.e., CH condition, Figure 3). Students were allowed to use hints without any limitations when they took the online test in the MH condition or CH condition. No points would be deducted, irrespective of choosing any hints.

Regardless of any conditions, the online test was delivered in the form of 30 multiple-choice questions, which were designed by an English expert. Each question was provided with three options, one of which was correct. If a wrong option was selected, five points would be reduced. Furthermore, the online test was implemented with a
10-inch android mobile device. This is owing to the fact that its display screen has a proper size. By doing so, such a mobile device not only demonstrates acceptable readability, but also offers the convenience of accessing information (Lan, Sung, Tan, Lin, & Chang, 2010; Huang, Wu, & Chen, 2012).

Figure 2. NH condition

Figure 3. CH condition

Empirical study

In this study, we used a quasi-experimental design to examine the impacts of cognitive styles on the use of hints in an online test. 46 individuals participated in this study. Participants were research students at a university in Taiwan and they volunteered to take part in this study. All participants had general English abilities but knew very few things about academic English.

Prior to conducting the experiment, the participants were given a series of academic English courses. During the experiment, they initially needed to take the Study Preference Questionnaire (SPQ). The SPQ includes two sets of 17 statements and the participants were asked to indicate the degree of agreement with either statement or to indicate no preferences. This study identified Holists and Serialists by using criteria suggested by Ford (1995): (a) if students agree with over half of the statements related to Holists, they are identified as Holists; (b) if students agree with over half of the statements related to Serialists, they are then considered as Serialists. Due to such simple criteria, the SPQ was chosen for the current study.

The results from the SPQ indicated that the sample consisted of 24 Holists and 22 Serialists. Subsequently, the participants were requested to interact with the aforementioned three versions of the online test. It took about 15 minutes to interact with each version. Furthermore, students needed to use each version to complete a task, which included 10 questions. How they completed the tasks was recorded in a log file, including the frequencies and time of using various hints and total time spent for interacting with the online test. The data recorded in the log file were applied to identify their learning behavior and they also needed to complete a formal test, which also consists of 20 multiple-choice questions. Finally, they were requested to fill out a questionnaire, which consisted of five opened questions, i.e., (1) the advantages of using the online test, (2) the disadvantages of using the online test, (3) English skills learnt from the online test, (4) the difficulty of using the online test and (5) the favorite type of hint. These five questions were applied to examine their responses to the online test.
Data analyses

Data analyses were conducted with an integrative approach, which included quantitative analyses, qualitative analyses and lag sequential analyses.

- **Quantitative Analyses**: A pair-t test was used to analyze differences between each condition and Pearson’s correlations were employed to analyze relationships between the hints used in each condition and the test scores. Furthermore, an independent t test was applied to identify differences between Holists and Serialists, in terms of learning performance and learning behavior. The former was measured based on their scores obtained from the tasks and tests, i.e., the task scores and the test score, and the time spent for completing the tasks, i.e., the task time. On the other hand, the latter was identified on the basis of the frequencies of using various hints, including Chinese Hints and Synonyms Hints.

- **Qualitative Analyses**: Two researchers who had strong background in qualitative methodologies and had interests in English learning were responsible for reading and coding responses to the open questions of the questionnaire based on procedures suggested by Merriam (1988). Subsequently, similar responses were clustered together and were given headings that best described the characteristics of the responses in each cluster, for which a frequency count was performed. Finally, the responses from Holists and those from Serialists were compared so that we could identify similarities and differences between them.

- **Lag Sequential Analyses**: This study involved multiple conditions and multiple hints, which increased the complexity of data. Thus, only using the aforementioned quantitative and qualitative analyses was not sufficient. To address this issue, the lag sequential analysis was also applied in this study, where sequential relationships of the behaviors were illustrated with visual diagrams so that behavior differences between Holists and Serialists could be clearly identified. The other benefit of using the lag sequential analysis is that it can determine how often one behavior was followed by the other (Lehmann-Willenbrock & Allen, 2014). Accordingly, we used the lag sequential analysis to discover which behavior could lead students to get a correct or wrong answer in this study. Such sequential relationships can help instructors identify an online test that can enhance students’ performance and facilitate students to choose suitable hints for themselves.

Results from quantitative analyses

Results from quantitative analyses are presented in this section, where we begin to describe overall tendencies and then move to depict the effects of cognitive styles on learning performance and learning behavior. Finally, relationships between the task scores and the test scores for each cognitive style group are presented.

Overall tendencies

In this section, we analyzed overall tendencies by identifying correlation between the task scores and the test scores and between the time spent for completing the tasks and the test scores for all students. The results indicated that students’ task scores were significantly related to their test scores ($r = .490$, $p < .05$). However, no significant correlations existed between the task time and the test scores.

Further to their performance in the tasks, we also investigated the relationships between the test scores and the frequencies of using various hints for all students. We found that the test scores were significantly associated with the frequencies of using synonym hints ($r = .523$, $p < .05$), instead of Chinese hints ($p > .05$). More specifically, students who more used synonym hints could obtain higher test scores. These results implied that the synonymy hints were more helpful to the students than Chinese hints, in terms of the test scores. This might be because synonymy hints could assist students to deliver their thoughts with multiple ways while Chinese hints could help them understand the Chinese meanings of vocabularies only. Thus, there is a need to encourage students to use synonymy hints, with which they could effectively express their ideas.

Learning performance

As mentioned before, students were provided with three conditions, i.e., the NH condition, MH condition and CH conditions. As shown in Table 1, Holists and Serialists obtained similar task scores in all conditions ($p > .05$). On the other hand, we also examined the effects of cognitive styles on the task time in each condition. Like
the task scores, there were no significant differences between Holists and Serialists, regardless of any conditions ($p > .05$). In particular, they spent the similar amount of time for completing the tasks in the NH condition, which did not provide any hints.

### Table 1. Performance in each condition

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Groups</th>
<th>Task scores Mean (SD)</th>
<th>Task time Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MH</td>
<td>Holists</td>
<td>88.75 (7.11)</td>
<td>176.42 (48.52)</td>
</tr>
<tr>
<td></td>
<td>Serialists</td>
<td>90.45 (5.22)</td>
<td>136.64 (51.34)</td>
</tr>
<tr>
<td>NH</td>
<td>Holists</td>
<td>77.92 (19.40)</td>
<td>193.42 (85.61)</td>
</tr>
<tr>
<td></td>
<td>Serialists</td>
<td>81.83 (11.02)</td>
<td>199.36 (87.92)</td>
</tr>
<tr>
<td>CH</td>
<td>Holists</td>
<td>85.42 (11.78)</td>
<td>179.08 (74.60)</td>
</tr>
<tr>
<td></td>
<td>Serialists</td>
<td>86.36 (9.77)</td>
<td>210.18 (86.76)</td>
</tr>
</tbody>
</table>

Furthermore, we analyzed how Holists and Serialists reacted to these three conditions. Regarding the task scores, students performed similarly in the three conditions, regardless of Holists ($p > .05$) or Serialists ($p > .05$). Regarding the task time (Figure 4), Holists spent similar amounts of time for completing the tasks in these three conditions ($p > .05$). Conversely, the task time that Serialists spent for each condition was different ($p < .05$). More specifically, Serialists in the NH condition significantly spent more time completing the tasks than those in the MH condition ($t = 3.33$, $p < .05$). Additionally, it significantly took more time for Serialists in the CH condition to complete the tasks than those in the MH condition ($t = 2.88$, $p < .05$). However, no significant difference was found between Serialists in the NH condition and those in the CH condition ($p > .05$). These findings suggested that the lack of hints might make Serialists spend much more time completing the tasks. In other words, Serialists more relied on the hints than Holists. Such results echoed those shown in Chen and Chang (2016), which revealed that Serialists needed additional support.

![Figure 4. The task time in each condition](Second)

### Learning behavior

Whether Holists and Serialists demonstrated different learning behavior in these three conditions is investigated in this section. The results indicated that Holists and Serialists behaved similarly in the CH condition ($p > .05$) and NH condition ($p > .05$). However, there was a significant difference between Holists ($Mean = 3.83$, $SD = 4.63$) and Serialists ($Mean = .64$, $SD = 2.11$) in the MH condition, in terms of the frequencies of using the hints (Figure 5). More specifically, Holists more frequently used the hints than Serialists ($t = 2.16$, $p < .05$). After analyzing the frequencies of using each type of hints, we found that such a significant difference majorly existed in synonym hints ($t = 2.36$, $p < .05$), instead of Chinese hints ($p > .05$). To be more specific, Holists ($Mean = 2.33$, $SD = 1.87$) more frequently used synonym hints than Serialists ($Mean = .27$, $SD = .04$). In other words, Serialists less frequently used hints, especially synonym hints (Figure 6).

On the other hand, the results from Pearson’s correlations indicated that no significant correlations existed between Holists’ task scores and the frequencies of using Chinese hints ($p > .05$) and synonym hints ($p > .05$).
However, Serialists’ task scores were related to the frequencies of using Chinese hints \((r = .048, p < .05)\) as well as those of using synonym hints \((r = .048, p < .05)\). These findings implied that Serialists might use Chinese hints and synonym hints in a relatively smart way. Accordingly, the more hints they used, the more task scores they got.

![Figure 5. The frequencies of using hints](image)

![Figure 6. The use of hints in the MH condition](image)

**Task scores vs. test scores**

As mentioned before, a significant correlation existed between students’ task scores and their test scores. In this section, we further analyzed such a correlation from a cognitive style perspective. The results indicated that no such significant correlation was found for Serialists while there were some significant correlations for Holists.

In general, Holists’ task scores in the three conditions was significantly related to their test scores \((r = .628, p < .05)\). After analyzing the performance of Holists in each condition, we found that such a correlation was majorly happened in the NH condition. More specifically, Holists’ task scores in the MH \((p > .05)\) and CH \((p > .05)\) conditions were not significantly related to their test scores but their task scores in the NH condition were significantly associated with their test scores \((r = .626, p < .05)\). The difference between the NH conditions and the remaining two conditions lied within the fact that the former did not offer any hints while the latter provided some hints. In other words, they could not get any additional support in the NH condition so they had to rely on their prior knowledge. This finding revealed that Holists could also make the best use of their prior knowledge to cope with the test, instead of depending on the tasks only. This finding suggested that they could not only use what they had learned from each task, but also take advantage of their prior knowledge. This might be because Holists took a comprehension learning style to view the tasks and their prior knowledge as a whole (Howie, 1995).
Results from qualitative analyses

This section presents the results from their responses to the opened questions of the questionnaire, which were analyzed with a qualitative approach. Both Holists and Serialists shared several similar responses while some differences also existed between them.

Similarities between Holists and Serialists

Holists and Serialists showed similar responses to the advantages and disadvantages of using the online test and enhanced English skills. The details are presented in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Similar Responses</th>
</tr>
</thead>
</table>
| Advantages of using the Online test | • Immediate feedback to know the answers  
• Lightweight and easy to take  
• No limitations for time and places  
• Simple and visual interface  
• Full of curiosity and freshness  
• No need to make hard copies |
| Disadvantages of using the Online test | • Uncomfortable to use a touch screen  
• Unable to answer questions non-sequentially  
• Unable to take the notes  
• No chance to correct wrong answers |
| Enhanced English skills | • Sentences  
• Prepositions  
• Grammatical issues  
• Sentence structure  
• Tenses of Verbs |

Differences between Holists and Serialists

On the other hand, differences also existed between Holists and Serialists. Firstly, they valued different types of hints. More specifically, Holists (N = 15, 63%) favored to use synonym hints while Serialists (N = 13, 59%) preferred to use Chinese hints. Furthermore, differences also existed in the aspects of the value of this online test and the difficulties that they met. Regarding the value of this online test, Holists paid attention to the whole test strategies. For instance, their thoughts were:

"It is valuable to me that the online test allows me to make a guess."
"The online test is useful because it facilitates me to read the questions carefully."
"I appreciate that the online test helps me use multiple test strategies."
"The online tests let me make effective use of resources to answer questions."

Conversely, Serialists emphasized on a single item provided by the online test, e.g.:

"The provision of hints is very helpful to me."
"I learn some new vocabularies from this online test."
"I learn how to use hints smartly to answer the question."
"I do not need to write the answer to each question"

Regarding the difficulties that they met, Holists focused on questions examined in the online test. For example, Holists thought:

"I feel difficult to understand the questions."
"I could only use limited time to answer the questions."
"Answering the questions needed to put heavy mental effort."

In contrast, Serialists were concerned with additional support, e.g.,

"There was a lack of an Introduction on how to take the test."
"I was not allowed to rotate the screen to answer the questions."
"I needed an alert before I submitted a wrong answer."
Results from lag sequential analysis

This section presents the results from the lag sequential analysis, including behavior frequency analysis, behavior sequential analysis and learning behavior patterns.

Behavior frequency analysis

The individual behaviors of the participants were coded using the coding scheme, which included five kinds of codes (Table 3). According to the rationale of the lag sequential analysis, participants’ behaviors were coded in the chronological order of their occurrences. For example, after using Synonymy Hint 1 (S1), Synonymy Hint 2 (S2) and the Chinese hint (C), then they will get a wrong answer (N) to the question; this series of behaviors was, thus, coded as S1S2CN. In total, there are 645 students’ behavior codes.

<table>
<thead>
<tr>
<th>Code</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Use Synonymy Hint 1</td>
</tr>
<tr>
<td>S2</td>
<td>Use Synonymy Hint 2</td>
</tr>
<tr>
<td>C</td>
<td>Use Chinese Hint</td>
</tr>
<tr>
<td>Y</td>
<td>Get Correct an Answer</td>
</tr>
<tr>
<td>N</td>
<td>Get Wrong an Answer</td>
</tr>
</tbody>
</table>

Behavior sequential analysis

Based on the frequency transition tables (Table 4 and Table 5), we examined whether the connection between each sequence reached statistical significance. The z-score value of each sequence was calculated to determine whether the continuity of each reached the level of significance and a z-value greater than +1.96 indicated that a sequence reached the level of significance ($p < .05$). In that case, the codes obtained from Holists and Serialists yielded the adjusted residuals tables (Tables 6 and 7). Furthermore we deduced the behavior-transfer diagrams of Holists and Serialists. Figure 7 and Figure 8 illustrate all sequences that have reached significance and the numerical values in the figures are the sequences’ z-scores and the arrow indicates the direction of transfer for each sequence.

<table>
<thead>
<tr>
<th>Holists</th>
<th>S1</th>
<th>S2</th>
<th>C</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>S2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Y</td>
<td>10</td>
<td>0</td>
<td>12</td>
<td>70</td>
<td>15</td>
</tr>
<tr>
<td>N</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>19</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Serialists</th>
<th>S1</th>
<th>S2</th>
<th>C</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>S2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>80</td>
<td>14</td>
</tr>
<tr>
<td>N</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Holists</th>
<th>C</th>
<th>Y</th>
<th>N</th>
<th>Serialists</th>
<th>C</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1</td>
<td>27</td>
<td>12</td>
<td>C</td>
<td>1</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Y</td>
<td>30</td>
<td>62</td>
<td>16</td>
<td>Y</td>
<td>22</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>N</td>
<td>5</td>
<td>24</td>
<td>6</td>
<td>N</td>
<td>1</td>
<td>22</td>
<td>7</td>
</tr>
</tbody>
</table>
Table 6. The sequential analysis of behavior in the MH condition

<table>
<thead>
<tr>
<th></th>
<th>Holists</th>
<th></th>
<th></th>
<th>Serialists</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
<td>S2</td>
<td>C</td>
<td></td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>S1</td>
<td>-1.19</td>
<td>7.44</td>
<td>-0.41</td>
<td>-0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>2.19</td>
<td>-0.77</td>
<td>1.05</td>
<td>0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>-0.41</td>
<td>1.05</td>
<td>-1.26</td>
<td>0.10</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>0.04</td>
<td>-2.57</td>
<td>0.49</td>
<td>0.13</td>
<td>-0.18</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-0.16</td>
<td>-0.38</td>
<td>-0.90</td>
<td>0.88</td>
<td>0.24</td>
<td></td>
</tr>
</tbody>
</table>

Serialists

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>C</th>
<th></th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>-0.16</td>
<td>-0.12</td>
<td>-0.23</td>
<td>0.44</td>
<td>-0.53</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>-0.12</td>
<td>-0.08</td>
<td>5.96</td>
<td>-0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>-0.23</td>
<td>-0.16</td>
<td>-0.33</td>
<td>0.63</td>
<td>-0.75</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>-0.39</td>
<td>0.31</td>
<td>0.04</td>
<td>-0.13</td>
<td>-0.38</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-0.53</td>
<td>-0.37</td>
<td>-0.75</td>
<td>0.43</td>
<td>0.62</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05.

Figure 7. The behavioral transition diagrams in the MH condition

Holists

Serialists

Figure 8. The behavioral transition diagrams in the CH condition

Learning behavior patterns

After comparing the behavioral diagram of Holists with that of Serialists, we found that Holists and Serialists demonstrated significant different behavior sequences. Regarding the MH condition, the significant sequence, S1 ↔ S2, was found for Holists, indicating that they used Synonymy Hint 1, then they would use Synonymy Hint 2, and subsequently they would use Synonymy Hint 1 again. On the other hand, the significant sequence, S2 → C, was found for Serialists, indicating that Serialists used Synonymy Hint 2 and then they would use the Chinese hint. In other words, Holists demonstrated iterative behavior while Serialists showed sequential behavior. Such findings implied that Holists relied on synonymy hints. However, Serialists did not depend on synonymy hints so much because they would also use Chinese hints. A difference between the synonymy hints and Chinese hints lied within the fact that the former belonged to indirect hints, which forced students to do the translation twice, while the latter pertained to direct hints, which allowed students to do the translation once. In other words, the latter was more directly relevant to their current task than the former. Accordingly, it seemed that Serialists tended to take an approach that was directly related to their task. These findings are coherent with those of Clewley, Chen, and Liu (2010), which indicated that Serialists preferred to only use the options that are relevant to their current tasks.
Regarding the CH condition, no significant sequence was found for Serialists. However, the significant sequence, C→ N, was found for Holists, indicating that Holists used the Chinese hints and then got a wrong answer. The findings implied that using Chinese hints had a possibility to make Holists get a wrong answer. Holists were suitable to have educational materials with rich information (Howie, 1995) but only Chinese hints were provided in the CH condition. Thus, the CH condition did not match with the needs of Holists. In brief, the aforementioned results from lag sequential analyses suggested that cognitive styles had great effects on student’s behavior patterns.

Discussions

Figure 9 presents a framework, which summarizes the findings of this study. As shown in this figure, cognitive styles have considerable influences on students’ learning patterns when they using the online test. In particular, they used hints very differently. Such behavior differences may be a reflection of differences in their cognitive styles (Wildemuth et al., 1998). The details are discussed below.

Holists took an iterative approach

The results from Holists’ qualitative responses to the opened questions of the questionnaire indicated that Holists valued multiple test strategies provided by the online test. This may be the reason why Holists repeated to use two types of synonymy hints. Due to such iterative behavior, the frequencies of Holists using the hints were higher than those of Serialists. This result was coherent with that of Clewley, Chen, and Liu (2010), which indicated that Holists preferred to have all of the available options. In other words, they prefer to collect large amounts of information (Howie, 1995), instead of focusing one a single item. This might be due to the fact that Holists tended to be relatively global (Mampadi, Chen, Ghinea, & Chen, 2011) so they took breadth-first paths. Hence, such a global approach made them demonstrate iterative behavior.

Serialists needed additional support

The results from quantitative data indicated that Serialists’ task scores were significantly related to the frequencies of using Chinese hints and those of using synonym hints. Thus, using Chinese hints and synonym hints was useful for Serialists to get higher task scores. Furthermore, Serialists in the NH condition and those in the CH condition significantly spent more time for completing the tasks than those in the MH condition. In other words, Serialists spent more time for completing the tasks if fewer hints or no hints were offered. These findings implied that Serialists needed additional support from hints. This might be owing to the fact that Serialists tended to take a suggested route (Clewley, Chen, & Liu, 2011) and the hints provided such a suggested route for Serialists. This might be the reason why Serialists relied on hints.

![Figure 9. The framework to summarize the findings](chart)

Figure 9. The framework to summarize the findings
Conclusions

This study aimed to investigate how cognitive styles affected learners’ reactions to the use of hints in the context of academic English, in terms of learning behavior and learning performance. To obtain deep understandings, a learning-analytics approach was applied in this study, including quantitative measurement, qualitative evaluation and lag sequential analyses. The results from the quantitative measurement indicated that the frequencies of using Chinese hints and synonym hints were significantly associated with Serialists’ task scores while such significant correlations were not found for Holists. The results from the qualitative evaluation showed that Holists favored to use synonym hints while Serialists preferred to use Chinese hints. The results from the lag sequential analyses suggested that Holists demonstrated iterative behavior while Serialists showed sequential behavior. In summary, cognitive styles have considerable influences on students’ learning patterns in the context of Academic English.

Thus, designers should consider how to develop personalized online tests that can accommodate the preferences of different cognitive style groups based on the findings of this study. However, this study has several limitations. Firstly, the sample is small so further works need to use a bigger sample to verify the findings presented in this study. Additionally, this study took into account differences between Holists and Serialists only. Therefore, future research should consider other dimensions of cognitive styles (e.g., Field Dependent and Field independent). By doing so, more comprehensive results could be obtained.

Acknowledgements

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan, for financial support (MOST 104-2511-S-008-008-MY3)

References


A Mobile Game-Based English Vocabulary Practice System Based on Portfolio Analysis

Ting-Ting Wu¹ and Yueh-Min Huang²

¹Graduate School of Technological and Vocational Education, National Yunlin University of Science and Technology, Yunlin, Taiwan // ²Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan // twu@yuntech.edu.tw // huang@mail.ncku.edu.tw

*Corresponding author

ABSTRACT

English learning has become a vital educational strategy in many non-English-speaking countries. Vocabulary is a critical element for language learners. Therefore, developing sufficient vocabulary knowledge enables effective communication. However, learning a foreign language is difficult and stressful. In addition, memorizing English vocabulary is often considered boring, and learners may lack motivation for learning activities. To increase English learning motivation and interest, this study constructed a mobile game-based English vocabulary practice system that entails selecting words according to textbook passages, a difficulty ratio, and learning portfolios. The learning activities involved in the system transform vocabulary learning from tedious memorization to game-based learning, thereby enhancing learners’ vocabulary memory and their familiarity with curriculum-related vocabulary through various multimedia. An experiment was conducted by dividing students into three groups that involved different vocabulary review methods. All students had similar English proficiency levels, and the course content and instructor were the same for all groups. This study statistically analyzed whether learners become familiar with vocabulary after playing the game. According to the analysis results, students who used the proposed system exhibited higher learning interest, attention, and learning effectiveness, as well as a sense of accomplishment and triumph, compared with other students.

Keywords

Game-based learning, English learning, Leaning portfolio, Lag Sequential Analyses

Introduction

In the era of globalization, English has become the common language of international communication (Spolsky & Shohamy, 1999). The importance of English is evident in international trade as well as leisure and entertainment. English learning has become a vital educational strategy and policy in numerous non-English-speaking countries. Vocabulary is particularly critical for language learners, and developing a sufficient vocabulary is necessary for achieving effective communication (Nation, 2001). Nation (2006) reported that in listening or speaking, a vocabulary of 6000–7000 words is necessary to achieve more than 98% comprehension. In addition, Laufer (2001) indicated that knowledge of vocabulary is highly correlated with writing and reading ability as well as academic performance. Jordan (2007) also suggested that difficulties in writing are mostly due to insufficient knowledge of vocabulary. Therefore, lack of sufficient knowledge of vocabulary hinders appropriate expression in writing. Oberg (2011) proposed that sufficient vocabulary is critical for academic achievement and learning effectiveness.

Learners of English as a second language (ESL) or English as a foreign language (EFL) generally consider language learning to be difficult and stressful (Turgut & İrgin, 2009). Moreover, learners typically consider that memorizing English vocabulary is a boring learning activity (Chen & Chung, 2008). This perception affects the learning interest and achievement of learners, which consequently diminishes their confidence in learning English. Therefore, instructional methods and strategies should be adjusted to increase the motivation and interest of learners to learn English (Chang, Liang, Yan, & Tseng, 2013; Jong, Lai, Hsia, Lin, & Lu, 2013).

Learning activities derived from instructional strategies influence the selection, acquisition, and construction of information for learners, which in turn affect the behaviors and thoughts expressed in their learning portfolios (Weinstein & Mayer, 1986). To achieve English vocabulary objectives, increase learning motivation and interest, and facilitate active learning and flow experience, digital game-based learning approaches can be used in an EFL learning environment (Ryu & Parsons, 2012).

The multimedia features of digital game-based learning are composed of text, voice, and images. Digital game-based learning can effectively enhance the attention, interest, creativity, and community relationships of students. Moreover, designing a series of rules and objectives in a digital game-based learning environment can enable achieving mental and physical satisfaction and insight; such satisfaction and insight can thus facilitate the realization of learning objectives (Burguillo, 2010). Games are typically characterized by curiosity, expectation, control, and interactive features, which can increase learners’ learning interest and intrinsic motivation (Konradt...
& Sulz, 2001; Lim, Nonis, & Hedberg, 2006). Learners are willing to overcome difficult challenges to gain a sense of achievement. Compared with conventional courses, digital game-based learning enables students to improve their memory of course content and engage in more critical thinking (Ke, 2008; Ke, 2014; Papastergiou, 2008). Digital game-based learning, which combines entertainment and education activities, has been researched in various academic fields (Liu, 2014; Liu, Cheng & Huang, 2011; Miller, Robertson, Hudson, & Shimi, 2012; Shih, Shih, Shih, Su, & Chuang, 2010; Vos, Van der Meijden, Denessen, 2011). Digital game-based learning not only reduces the English learning anxiety experienced by ESL students, but also facilitates such students in cultivating interests toward learning; this thus increases the students’ motivation to learn English. For example, Uzun, Cetinavci, Korkmaz, and Salihoglu (2013) constructed a game to help EFL university students who were enrolled in basic English courses to practice English vocabulary. Their results showed that most students had positive attitudes and satisfaction toward game-based learning in the classroom; this finding demonstrates that game-based learning systems can support EFL instruction. Chou (2014) introduced game-based learning to students enrolled in elementary school English courses and revealed that game-based learning could effectively help learners to understand the meaning of English vocabulary and the content of instruction; this could increase the motivation to learn English, enhance English learning performance, and make English learning easier and more interesting. In addition, Chen and Yang (2013) used the video game BONE to teach EFL students, as well as to understand the views of learners and benefits gained through the introduction of games. They found that the students generally believed that game-based learning could effectively improve their English learning skills and motivation, and that most of the students enjoyed learning English through games. Moreover, the introduction of the video game BONE effectively improved the students’ English listening, reading, and vocabulary techniques and skills.

Because of the gradual development of wireless technology and mobile devices, topics related to mobile-assisted language learning (MALL) have received considerable attention. Many studies have incorporated portability, interactivity, connectivity, privacy, and real-time features into language-learning courses (Chen & Li, 2009; El-Bishouty, Ogata & Yano, 2007; Petersen, Markiewicz, & Bjornebekk, 2009; Wu, Sung, Huang, Yang & Yang, 2011). Chang, Tseng, Liang, and Yan (2013) constructed an English learning system for PDA devices used by 125 high school students; they explored the students’ continued use of the system and perspectives regarding the benefits of the system. Their results revealed that most students had a positive view toward using mobile devices for English learning and believed that the system was beneficial for English learning. Furthermore, Hsu (2013) investigated the perceptions of EFL students in different countries toward MALL and reported that most learners believed that mobile technology is a useful and practical tool for language learning. Combining MALL with game-based learning also creates a revolutionary learning model and environment for language learning. Yu and Guan (2013) constructed an English vocabulary passing game on smartphones; they demonstrated the convenience and practicality of the devices, which enabled learning to become a part of life. Game-based learning can effectively increase interest in and motivation for engaging in English learning, which consequently enhances English vocabulary knowledge. Smith et al. (2013) constructed a game-based English vocabulary-learning system on e-books and experimentally demonstrated that their interactive e-book game design increased the motivation for engaging in English learning. Real-time learning tools and related assistant functions can effectively promote students’ English-reading ability, thereby increasing the benefits of English learning.

According to the aforementioned advantages, this study constructed a mobile game-based English vocabulary practice system on the basis of learning objectives, learning needs, and the learning scope. In the vocabulary presentation game, words are selected according to textbook passages, a difficulty ratio, and learning portfolios. This system could enable learners to achieve autonomous learning and self-planning according to the curricular objectives and needs; moreover, the teacher used instructional objectives and the learning scope to set the game content to meet curricular objectives. Regarding system development and construction, the game imitates a type of block-clearing game that is popular among young students. The game converts the process of vocabulary memorization, which is considered by students to be an uninteresting activity, into a game-based learning activity so that learners can become proficient in English vocabulary. When learners finish reading an article, they can play spelling games with diverse media effects. Learners can familiarize themselves with vocabulary by playing the game, which can even serve as a formative assessment. The purpose of this study was to provide learners with a relaxed and pleasant English vocabulary-learning environment that can reduce their learning anxiety and stress and increase their learning interest and initiative. Learners can experience improving achievement in learning English vocabulary, and this improvement is engendered by an inadvertent increase in the learning time and concentration.
Mobile game-based English vocabulary practice system

System framework

For meeting the instructional objectives of English learning and the digital game-based instructional strategy, and for promoting the introduction of mobile devices, this study provided a game-based learning environment to immerse learners in game-based learning and improve their learning experience. This study used the input–process–outcome game model implemented by Garris, Ahlers, and Driskell (2002) to design a mobile game-based English vocabulary practice system on the basis of learning procedures and needs; students using this can be immersed in the game and fully enjoy the learning process.

A three-tier model was used for constructing and developing the framework of the proposed system (Figure 1). The three-tier framework comprises the presentation tier, application tier, and data tier. The presentation tier primarily processes data presentation, the operational interface and functions, and necessary data exchange with the application tier. The application tier is responsible for extensive data processing and computational procedures for reducing the computational load imposed on mobile devices with limited resources; this tier serves as the middle tier between the presentation tier and data tier. The data tier is the database server and functions as a centralized data storage and management system to provide high-efficiency data access functions.

On the learner end, mobile phones, tablet computers, and other widely used smart mobile devices serve as game-based learning devices. The application server of the game system can be connected to the web through Wi-Fi wireless Internet or 3G (HSPA), 4G (LTE), or other mobile communication networks. Data transfer between the two ends can be achieved using either a socket client/server or HTTP request/response model. The independent database server can be placed on an intranet to increase data security, and the application server connects the Internet and intranet.

System interface and functions

The proposed game-based assistive English vocabulary practice system was developed with the objective of providing a software tool that can be used by students during class or after class for review, practice, and repeated learning; its development was not aimed at replacing official classroom instruction. The game imitates a block-clearing game that is popular among students, and it supports English vocabulary, sentence, and listening practice. The learner plays the game with selected sentences, and the goal is to clear continuous letters that constitute English vocabulary words rather than blocks of the same color as in similar games. In addition, it provides a model that
enables learners to practice listening to vocabulary words and sentences; specifically, learners clear English vocabulary words or sentences after listening to them recited aloud. Words and sentences are selected according to three factors:

- **Textbook content:** The system automatically selects vocabulary words or sentences from a textbook article for gameplay; any word or sentence may be selected. Learners can use this method to review and familiarize themselves with the vocabulary words and sentences in the textbook.

- **Difficulty ratio:** The frequency ratio of the appearance of each vocabulary word or sentence is related to its difficulty. The frequency ratio of sentence appearance is based on a positive or negative correlation with difficulty. If positive, the appearance frequency is higher for more difficult sentences; if negative, the appearance frequency is higher for easier sentences. The difficulty indices are based on the two definitions of the Flesch reading ease readability formula and GEPT word lists, which have undergone adjustments and revisions. The formula applied to the proposed system considers two parameters in the Flesch formula: the number of syllables per vocabulary word and the length of the sentence, as well as more than 8000 vocabulary words at the GEPT elementary, intermediate, and high-intermediate levels. In the formula, \( r_{DM}^v(v) \) is the ratio of vocabulary word occurrences, \( NoS(v) \) is the number of syllables in vocabulary word \( v \), \( GEPT(v) \) is the level of vocabulary word \( v \) (elementary = 1, intermediate = 2, and high intermediate = 3), and \( V \) is the set of vocabulary words in the textbook.

\[
\begin{align*}
r^v(v) &= \frac{1}{2} \left( \frac{NoS(v)}{\sum_{v' \in V} NoS(v')} + \frac{GEPT(v)}{\sum_{v' \in V} GEPT(v')} \right) \\
r^s(s) &= \frac{1}{2} \left( \frac{LoS(s)}{\sum_{s' \in S} LoS(s')} + \frac{r^s(s)}{\sum_{s' \in S} r^s(s')} \right)
\end{align*}
\]

The calculation of the sentence occurrence ratio \( r_{DM}(s) \) entails considering the length of sentences, the syllables of all vocabulary words constituting the sentence, and the GEPT level. In the following formula, \( LoS(s) \) represents the length of an \( s \) sentence (number of vocabulary words in the sentence), \( S \) represents a set of sentences in the textbook, and \( V_i \) represents a vocabulary set constituting the \( s \) sentence.

\[
r_{DM}(s) = \left\{ \begin{array}{ll} r(s), & \text{positive correlation} \\
1 - r(s), & \text{negative correlation} \end{array} \right.
\]

- **Learning portfolio:** The system extracts and analyzes a student’s previous learning records from the learning portfolio database. After analysis, sentences that should have been cleared but were not, or sentences that were incorrectly or unsuccessfully cleared, are selected from the learner’s previous game. This method can prevent the learner from always clearing shorter or easier sentences to accumulate points. The method can also prevent the game from unintentionally omitting vocabulary words; thus, learners have more opportunities to improve by reviewing and becoming more accustomed to unfamiliar vocabulary. Therefore, the system applies the binary exponential back-off algorithm to calculate the occurrence ratio in the next game for sentences that have not been successfully cleared. Words and sentences that have been missed more times exhibit a higher occurrence ratio. The occurrence ratio is expressed as

\[
\forall v \in \hat{V} = \bigcup_{i=1}^{n} \left( V_i - F_i^V \right) \\
t(v) = \sum_{i=1}^{n} f(v) \\
where f(v) = \begin{cases} 
1, & v \notin \hat{V} \\
0, & \text{otherwise} 
\end{cases} \\
r_{PM}(v) = \frac{2^t(v)}{\sum_{v' \in \hat{V}} 2^t(v')}
\]

In the formula, \( V_i \) is the set of selected words in the \( i \)th round of the game, \( F_i^V \) is the set of cleared words in the \( i \)th round, \( n \) is the number of the games previously played by the learner, \( \hat{V} \) is the vocabulary words that the learner
could not successfully clear in the previous game, \( t(v) \) is the number of vocabulary words \( v \) that should have been cleared in the previous game, and \( f(v) \) is a function representing whether a vocabulary word has been successfully cleared in a single-round game. The sentence occurrence ratio is calculated in the same manner, as follows:

\[
\pi_{FM}(s) = \frac{2^{t(v)}}{\sum_{s' \in \mathcal{S}} 2^{t(s')}}
\]

Figure 2 presents the operational interface. Most block-clearing games comprise connected square icons, and each letter can be connected to only four adjacent letters. By contrast, the proposed game system contains contiguous hexagons, and this enables each letter to be connected to six adjacent letters. The learner forms a vocabulary word by swiping his or her finger to connect letters. The corresponding letters disappear, causing the adjacent letters to change position. Subsequently, the learner can find another vocabulary word by connecting the adjacent letters. Moreover, the game system can produce questions by using four models. (1) English vocabulary model: The system shows English vocabulary words, and the learner finds the words and clears them. (2) English sentence model: The system shows English sentences, and the learner finds the vocabulary words constituting the sentence and clears them. (3) Chinese word model: The system shows a Chinese word, and the learner finds the corresponding English vocabulary word and clears it. (4) Chinese sentence model: The system shows a Chinese sentence, and the learner finds the corresponding English sentence and clears it.

Figure 2. Interface of the mobile game-based English vocabulary practice system

Research design

Participants

The participants in this study were first-year students from three classes in the department of information management at a private university in Taiwan; the students were enrolled in a basic English course. The participants were divided into three groups on the basis of their class: Group T, comprising 30 (18 male and 12 female) students; Group R, comprising 32 (20 male and 12 female) students; and Group G, comprising 32 (18 male and 14 female) students. The students in Group T learned English in the classroom by engaging in only paper-based tasks instructed through traditional lecture-style instructional methods. The students in Group E also learned English in the classroom through traditional instructional methods, and half the class period each week was allocated to vocabulary review. The students in Group G received the same instructional strategy as those in Group E did; however, the students in Group G used the mobile game-based English vocabulary practice system for vocabulary review, which was allocated half the class period. Because of the university admission mechanism (entrance exam) in Taiwan, freshmen in the same department have relatively the same English proficiency level. Therefore, to ensure that the students in the three groups received the same course content, the same instructor taught all three
courses in this study; moreover, the students in the three groups used the same instructional material database for the reading and learning processes.

Experimental procedures

The basic English course was conducted for two class periods per week in an 18-week semester. To meet the instructional objectives stipulated by the school, all the students in the three groups first underwent lecture-style instruction from Weeks 1 to 8. Midterm examinations were administered in Week 9; this study used this opportunity to conduct a pretest of the experiment. The planned experimental activities began after the midterm examinations. During Week 10, the English teacher used half the class period to explain the procedures and objectives of the experimental activity to the students in Group G and allowed them to operate and practice using the system. Thus, the students familiarized themselves with the system’s operational interface and environment. Between Weeks 11 and 17, the basic English course was conducted for two class periods per week. The students in Group T learned English through traditional lecture-style instruction. The students in Group R also learned English in the classroom through traditional lecture-style instruction, but the second half of the class period was allocated to English vocabulary review. The students in Group G learned English in the classroom through traditional lecture-style instruction; however, during English vocabulary review in the second half of the class period, they used the proposed mobile game-based English vocabulary practice system for practice. All operating behaviors were recorded in the back-end database. The introduction of the game improved the students’ learning motivation, intent, and interest; thus, they achieved active and independent learning, which consequently promoted the effectiveness of English learning. During the learning activity, the instructor and technicians helped the students to troubleshoot any problems with the learning activity and system use. Figure 3 illustrates a flowchart of the experimental procedure. Week 18 was the final examination week. In accordance with the school schedule, this study used the final examination as the posttest for this experiment.

Assessment tools

Learning achievement

Learning achievement can serve as a basis for performance evaluation to reflect the advantages and disadvantages of an instructional design and instructional strategy. To understand whether the mobile game-based English vocabulary practice system increases English proficiency and understanding and whether it benefits English vocabulary learning, this study statistically analyzed the pretest and posttest results to evaluate learning achievement. In addition to exploring differences in learning achievement associated with different learning methods, this study explored whether the proposed system promotes students’ English vocabulary ability.


Learning behaviors

Leaning behaviors are student reactions corresponding to learning contexts, particularly those caused by external environmental stimuli. Learning situations can be understood through learning behaviors for providing appropriate feedback to improve learning effectiveness (Politzer, 1983). To understand the learners’ operation and interaction, determine the relationships and sequences of their learning behaviors, and identify the frequency of and procedure for operating the system, this study encoded, defined, and classified the recorded learning portfolios. Subsequently, lag sequential analysis (LSA), which enables exploring and summarizing the cross-dependencies occurring in complex interactive sequences of behavior (Bakeman & Gottman, 1997; Hou, Chang & Sung, 2010), was adopted for the modularized evaluation of the relationships among and sequences and frequency of learning behaviors. The classification items were as follows: English vocabulary model (EV), English sentence model (ES), Chinese word model (CW), Chinese sentence model (CS), use assistance tool (A), checking rank (R), and pronunciation (P).

Research results

Analysis of learning effectiveness

This study evaluated learning effectiveness by analyzing the pretest and posttest scores of the students in Groups T, R, and G. Analysis of variance (ANOVA) and the general linear model were used for statistically analyzing the data. Descriptive statistics were used to explore the effects of the introduction of the mobile game-based English vocabulary practice system on the students’ performance in English vocabulary learning.

Table 1 presents the ANOVA results for the English performance of the three groups of students before the experimental instruction process. The results revealed no significant difference in the English learning performance of the three groups before the experimental instruction process ($F = .418, p > .05$, $\eta^2 = 0.79$).

<table>
<thead>
<tr>
<th>Table 1. ANOVA results of pretests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Group T</td>
</tr>
<tr>
<td>Group R</td>
</tr>
<tr>
<td>Group G</td>
</tr>
</tbody>
</table>

Table 2 lists the posttest results for Groups T, R, and G, obtained using ANOVA. The results revealed significant differences in the posttest performance of the three groups ($F = 5.38, p < .05$, $\eta^2 = 0.863$). Furthermore, the average scores of the groups showed that the Group R students outperformed the Group T students in learning achievement, because half the class period was allocated for English vocabulary review. Among the Group G students, introducing the game increased learning motivation and learning interest. Therefore, the performance of Group G was higher than that of Group R. Furthermore, according to the results of Scheffe post hoc comparisons, the learning achievement of the three groups was ranked as follows: Group T < Group R < Group G. In addition, the general linear model was used to evaluate the strength of correlations between different instructional strategies and learning achievement. According to the criteria of Cohen (1988) for determining relevance, a moderate correlation was observed between different instructional strategies and learning achievement ($\omega^2 = .086$).

<table>
<thead>
<tr>
<th>Table 2. ANOVA results of posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Group T</td>
</tr>
<tr>
<td>Group R</td>
</tr>
<tr>
<td>Group G</td>
</tr>
</tbody>
</table>

Note. *p < .05.

This study also used a dependent sample $t$ test to explore the differences between pretest and posttest values in the three groups. As presented in Table 3, the results indicated no significant difference between the pretest and posttest values in the students in Group T ($t = .884, p > .05, d = .72$). By contrast, a significant difference was observed in the students in Group R ($t = -3.892, p < .05, d = 0.69$). Furthermore, the mean pretest and posttest values indicated that the performance of the students in Group R improved slightly; this improvement was because they engaged in English vocabulary review for the second half of the class period per week. In addition, a significant difference was observed between the pretest and posttest values in the students in Group G ($t = -7.081, p < .05, d = 0.91$). Thus, reviewing English vocabulary by using the proposed system reduced the students’ anxiety and
obstructions regarding English learning. The game based-review system improved the students’ familiarity with and understanding of vocabulary, and this thus enhanced their English-learning performance.

Analysis of learning behavior

The behaviors of the students in Group G associated with the use of the mobile game-based English vocabulary practice system were recorded in the back-end database. Portfolio data were analyzed to explore the sequence and frequency of the students’ operation and interaction. First, the learning portfolios in the database were arranged according to the classification items. Subsequently, LSA was applied to evaluate the sequence and frequency of all learning behaviors. The assessment results for Group G are listed in Table 4.

Table 4. LSA assessment results for Group G

<table>
<thead>
<tr>
<th></th>
<th>EV</th>
<th>ES</th>
<th>CW</th>
<th>CS</th>
<th>A</th>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV</td>
<td>1573</td>
<td>1362</td>
<td>1423</td>
<td>876</td>
<td>583</td>
<td>792</td>
<td>249</td>
</tr>
<tr>
<td>ES</td>
<td>1482</td>
<td>1067</td>
<td>1395</td>
<td>681</td>
<td>768</td>
<td>502</td>
<td>192</td>
</tr>
<tr>
<td>CW</td>
<td>1490</td>
<td>983</td>
<td>1408</td>
<td>517</td>
<td>308</td>
<td>638</td>
<td>57</td>
</tr>
<tr>
<td>CS</td>
<td>736</td>
<td>495</td>
<td>632</td>
<td>487</td>
<td>397</td>
<td>248</td>
<td>33</td>
</tr>
<tr>
<td>A</td>
<td>579</td>
<td>757</td>
<td>298</td>
<td>370</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R</td>
<td>1097</td>
<td>848</td>
<td>907</td>
<td>430</td>
<td>0</td>
<td>59</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>231</td>
<td>138</td>
<td>49</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
</tbody>
</table>

In Table 4, the value in each cell with row items \( x \in \{ \text{EV}, \text{ES}, \text{CW}, \text{CS}, \text{A}, \text{R}, \text{P} \} \) and column item \( y \in \{ \text{EV}, \text{ES}, \text{CW}, \text{CS}, \text{A}, \text{R}, \text{P} \} \) represents behavior \( y \) occurring immediately after behavior \( x \). Finally, the \( z \) score was adopted for the analysis and computation according to the LSA results (Allison & Liker, 1982). The analytic results are presented in Table 5.

Table 5. Group G \( z \) score results

<table>
<thead>
<tr>
<th></th>
<th>EV</th>
<th>ES</th>
<th>CW</th>
<th>CS</th>
<th>A</th>
<th>R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>EV</td>
<td>10.83</td>
<td>7.38</td>
<td>8.85</td>
<td>5.23</td>
<td>2.16</td>
<td>4.52</td>
<td>-2.75</td>
</tr>
<tr>
<td>ES</td>
<td>9.57</td>
<td>6.46</td>
<td>7.93</td>
<td>3.02</td>
<td>4.07</td>
<td>1.42</td>
<td>-3.89</td>
</tr>
<tr>
<td>CW</td>
<td>9.74</td>
<td>6.03</td>
<td>8.28</td>
<td>1.76</td>
<td>-1.48</td>
<td>2.85</td>
<td>-5.85</td>
</tr>
<tr>
<td>CS</td>
<td>3.63</td>
<td>1.17</td>
<td>2.38</td>
<td>0.83</td>
<td>-0.13</td>
<td>-2.96</td>
<td>-6.98</td>
</tr>
<tr>
<td>A</td>
<td>2.01</td>
<td>3.89</td>
<td>-2.09</td>
<td>-0.57</td>
<td>-8.03</td>
<td>-8.79</td>
<td>-9.21</td>
</tr>
<tr>
<td>R</td>
<td>6.92</td>
<td>4.98</td>
<td>5.92</td>
<td>0.31</td>
<td>-8.84</td>
<td>-5.37</td>
<td>-9.38</td>
</tr>
<tr>
<td>P</td>
<td>-3.04</td>
<td>-4.68</td>
<td>-6.21</td>
<td>-7.17</td>
<td>-10.02</td>
<td>-10.32</td>
<td>-8.58</td>
</tr>
</tbody>
</table>

The analysis value was greater than 1.96, indicating a significant relevance and correlation between learning behaviors. To facilitate the research analysis, the information in Table 5 was converted into diagrams of behavioral relationships, as shown in Figure 4.

The line thickness represents the strength of the relationship between two behaviors. When the students used the proposed system, they could switch the model to experience a different practice method. According to Figure 4, the English vocabulary model was the most frequently used model, followed by the Chinese word model and the English sentence model; the least frequently used model was the Chinese sentence model. In addition, the students exhibited more active behaviors toward the English vocabulary and Chinese word models than the other models during the practice process. The results indicated that the students frequently switched between the English vocabulary and Chinese word models of the system. Moreover, the students frequently reviewed the ranking when they used the English vocabulary model and Chinese word model, because they needed to determine only one English vocabulary word and clear it. These two models were simpler than the other models. Therefore, the
students could quickly achieve the game levels, which promoted a mentality of competitiveness. Furthermore, because the English vocabulary model and English sentence model prompted English questions, the students naturally wanted to determine the words quickly and clear them. Therefore, the assistance tools were often used, especially for English sentences. Moreover, each game level has a time limit that students must complete before time-out. Although the system provides a pronunciation function, it was rarely used because of the pressure of the time-out.

Figure 4. Behavior pattern of Group G

Discussion

This study constructed a type of block-clearing game on the basis of learning objectives, learning needs, and the learning scope; this game is currently popular among students. This English vocabulary practice system entails the use of the game-based learning method and textbook content, difficulty ratios, and learning portfolios; thus, it enables students to practice and review vocabulary in a relaxed and enjoyable learning environment. The statistical analysis results regarding the three teaching strategies show that the students who received traditional lecture-style instruction passively accepted knowledge. Thus, no significant difference was observed in their learning effectiveness. By contrast, the learning effectiveness of the students in Groups R and G improved significantly. Our results are consistent with those of previous studies that have demonstrated that the percentage of forgotten words was highest 20 minutes after the students acquired knowledge. However, the proposed system facilitates remembering forgotten words, and it can effectively strengthen memory and learning ability and reduce learning stress after adequate memorization, review, and practice (Runquist, 1983; Miao, 2008; Wang, Liu, & Li, 2014). Furthermore, the posttest scores of Groups R and G reveal that Group G exhibited improved learning effectiveness; this finding indicates that the game-based learning method increased the students’ attention and interest. An interactive and diverse virtual environment as well as appropriate instructional material design and construction can provide students with an environment conducive to continual practice and review; this can effectively improve the students’ ability to face failure and setbacks as well as, in addition to enhancing their confidence and judgment (Ke, 2008, 2014; Papastergiou, 2008; Burguillo, 2010). The system’s question mechanisms repetitively receive information to facilitate students in increasing their familiarity with the system and enhancing their memory. Thus, the students in Group G exhibited the highest improvement in posttest performance (Ebbinghaus, 1964; Mayer, 1983; Mayer, 2001; Mayer, 2003; Churchill & Hedberg, 2008; Diependaele, Lemhöfer & Brysbaert, 2013).

Regarding learning behaviors, this study analyzed the learners’ operation of and interaction with the proposed system to understand the relationships and sequence of their learning behaviors. The statistical results show that the English vocabulary model and Chinese word model were frequently used during English vocabulary review, and the students demonstrated a more competitive and comparative mentality when they used these two models. According to the teacher’s observation, the students exhibited high learning interest and attention when they used the English vocabulary model and Chinese word model; this can be attributed to the models’ ease of operation. Consequently, it provided the students with a sense of accomplishment and triumph. As reported in previous studies, digital game-based learning is characterized by numerous features such as triumph, conflict, competition, and challenge; all of these features can immerse students in the learning process and thus prompt them to take the initiative to acquire knowledge and resolve problems, which can consequently engender active learning and self-
learning (Burguillo, 2010; Hainey, Connolly, Stansfield & Boyle, 2011; Jong, Lai, Hsia, Lin, & Lu, 2013; Prensky, 2001). Therefore, to immerse students in the learning process, instructors can incorporate some of the mentioned features in designing a game-based learning system. In addition, the students in the present study frequently utilized the assistance tools to accomplish the levels in the English vocabulary model and English sentence model. According to the students’ feedback after the experiment, the game system prompted English questions, which enabled them to find the answers and clear the words without translation. However, the time constraint made them feel anxious. Therefore, the assistance tools were used frequently to release anxiety under time pressure. Previous studies have reported that moderate anxiety can enhance student motivation and stress resistance (Chapell et al., 2005; Young, 1991). Consequently, in the learning process, students can be given appropriate pressure for enhancing their learning momentum.

**Conclusion and future works**

This study constructed a mobile game-based English vocabulary practice system on the basis of digital game-based learning. The textbook content, difficulty ratio, and learning portfolio were considered in selecting vocabulary. The learners enjoyed the game, and their motivation for and interest in English learning as well as their learning achievement also increased significantly. The behavior analysis results show that the introduction of the proposed system to students enhanced their competitive and comparative mentality, as well as their satisfaction and positive attitudes. Personal satisfaction is a crucial factor for the continuation of motivation (Keller, 1999; Small & Gluck, 1994; Lee, 2000), which is the guiding force behind behaviors and strength (Gagné, Yekovich & Yekovich, 1993). Therefore, higher learning motivation instills a sense of competence, self-efficacy, positive emotions, and expectations, which improve learning achievement (Cameron & Pierce, 1994; Elliott, & Dweck, 1988; Moos, 2014; Duffy & Azevedo, 2015). This study suggests that future studies establish an instructional design for English review that involves characteristics such as triumph, conflict, competition, and challenge, as well as similar question-writing strategies, to help students practice and review English vocabulary.

To meet the course schedule of the students in this study, the proposed system was introduced only after the midterm examinations. To achieve more in-depth analysis and exploration in the future, the proposed system can be introduced for the whole semester after sufficient analysis and discussion with instructors. Moreover, this study analyzed learning achievement and learning behavior. Future studies may analyze and explore cognitive and learning styles to understand students’ learning processes in detail when they use the system.

**Acknowledgements**

The research reported in this paper has been supported in part by the Ministry of Science and Technology (MOST) in Taiwan under the research project number MOST 103-2511-S-224-004-MY3, MOST 104-2511-S-224-003-MY3, and MOST 105-2628-S-224-001-MY3.

**References**


Retraction of - Volume 19, Issue 4, pp. 190-202

This article has been retracted at the request of the authors.

Due to highly unethical practices, which include data manipulation and falsification of results in this research, this article has been retracted. One of the conditions of the submission of a paper for publication is that authors declare explicitly that their work is original with the research results being conclusive from genuine research data. As such this article represents a severe abuse of the scientific publishing system. The editorial board takes a firm view on this matter and apologies are offered to readers of the journal that this was not detected during the submission process.