Educational Technology & Society
An International Journal

Aims and Scope

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

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

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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society.

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


Founding Editor
Kinshuk, 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; Yum 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; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Rob Koper, Open University of the Netherlands, The Netherlands; Jimmy Ho Man Lee, The Chinese University of Hong Kong, Hong Kong; Rudy Leleuiche, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McGray, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Vaxjo 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; Toshih Okamato, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J. Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Ching-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 Fominich, K. E. Tsolovskiy Russian State Tech Univ, Russia; Olga S Fomichova, Studio “Culture, Ecology, and Foreign Languages,” Russia; Piet Koomers, University of Twente, The Netherlands; Chul-Hwan Lee, Inchon National University of Education, Korea; Brent Muirhead, University of Phoenix Online, USA; Erkki Sutinen, University of Joensuu, Finland; Vladimir Uskov, Bradley University, USA.

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

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

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

Publisher
National Sun Yat-sen University, Taiwan

Editorial Office
c/o Dr Nian-Shing Chen, Department of Information Management, National Sun Yat-sen University, 70, Lien-Hai Rd, Kaohsiung, 80424, Taiwan.
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.E.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 4  2017**

## Table of contents

### Full Length Articles

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Middle School Students’ Mathematics Knowledge Retention: Online or Face-to-Face Environments</td>
<td>Clayton M. Edwards, Audrey C. Rule and Robert M. Boody</td>
<td>1–10</td>
</tr>
<tr>
<td>Recommending Learning Activities in Social Network Using Data Mining Algorithms</td>
<td>Lamia Mahbane</td>
<td>11–23</td>
</tr>
<tr>
<td>Analyzing Pauses in Computer-Assisted EFL Writing—A Computer-Keystroke-Log Perspective</td>
<td>Cuiqin Xu and Yan Qi</td>
<td>24–34</td>
</tr>
<tr>
<td>The Effect of Socially Shared Regulation Approach on Learning Performance in Computer-Supported Collaborative Learning</td>
<td>Lanqin Zheng, Xin Li and Ronghui Huang</td>
<td>35–46</td>
</tr>
<tr>
<td>Do Focused Self-Explanation Prompts Overcome Seductive Details? A Multimedia Study</td>
<td>Zhe Wang and Olusola Adesope</td>
<td>47–57</td>
</tr>
<tr>
<td>Students’ Reactions to Different Levels of Game Scenarios: A Cognitive Style Approach</td>
<td>Zhi-Hong Chen, Sherry Y. Chen and Chih-Hao Chien</td>
<td>69–77</td>
</tr>
<tr>
<td>Flow Experience and Educational Effectiveness of Teaching Informatics using AR</td>
<td>Stefanos Giassirantis and Loizos Sofos</td>
<td>78–88</td>
</tr>
<tr>
<td>Heuristics and Web Skills Acquisition in Open Learning Environments</td>
<td>Daniel Dominguez Figaredo</td>
<td>102–111</td>
</tr>
<tr>
<td>Effects of Attention Cueing on Learning Speech Organ Operation through Mobile Phones</td>
<td>Hui-Yu Yang</td>
<td>112–125</td>
</tr>
<tr>
<td>Effects of Self-explanation and Game-reward on Sixth Graders’ Algebra Variable Learning</td>
<td>Hong-Zheng San Lin and Guey-Fa Chiou</td>
<td>126–137</td>
</tr>
<tr>
<td>The Influence of a Pedagogical Agent on Learners’ Cognitive Load</td>
<td>Noah L. Schroeder</td>
<td>138–147</td>
</tr>
<tr>
<td>Development and Usability Test of an e-Learning Tool for Engineering Graduates to Develop Academic Writing in English: A Case Study</td>
<td>Chih-Chung Lin, Gi-Zen Liu and Tsong-I Wang</td>
<td>148–161</td>
</tr>
</tbody>
</table>

### Guest Editorial

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fostering Deep Learning in Problem-Solving Contexts with the Support of Technology</td>
<td>Minhong Wang, Sharon Derry and Xun Ge</td>
<td>162–165</td>
</tr>
</tbody>
</table>

### Special Issue Articles

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoXPT: Designing for Deeper Learning through Experimentation in an Immersive Virtual Ecosystem</td>
<td>Chris Dede, Tina A. Grotzer, Amy Kamarainen and Shari Metcalf</td>
<td>166–178</td>
</tr>
<tr>
<td>Design of a Three-Dimensional Cognitive Mapping Approach to Support Inquiry Learning</td>
<td>Juanjuan Chen, Minhong Wang, Chris Dede and Tina A. Grotzer</td>
<td>191–204</td>
</tr>
<tr>
<td>Leveraging Students’ Knowledge to Adapt Science Curricula to Local Context</td>
<td>Lani M. Minshew, Kelly J. Barber-Lester, Sharon J. Derry and Janice L. Anderson</td>
<td>205–218</td>
</tr>
</tbody>
</table>
Moving Apart and Coming Together: Discourse, Engagement, and Deep Learning
Andrea S. Gomoll, Cindy E. Hmelo-Silver, Erin Tolar, Selma Šabanović and Matthew Francisco
219–232

Deep Learning towards Expertise Development in a Visualization-based Learning Environment
Bei Yuan, Minhong Wang, Andre W. Kushniruk and Jun Peng
233–246

Deep and Surface Processing of Instructor’s Feedback in an Online Course
Kun Huang, Xue Ge and Victor Law
247–260

Investigating the Effects of Authentic Activities on Foreign Language Learning: A Design-based Research Approach
Ideniz Ozverir, Ulker Vanci Osam and Jan Herrington
261–274

Visualizing the Complex Process for Deep Learning with an Authentic Programming Project
Jun Peng, Minhong Wang and Demetrios Sampson
275–287

Can Students Identify the Relevant Information to Solve a Problem?
Lishan Zhang, Shengquan Yu, Baoping Li and Jing Wang
288–299

An Eye Tracking Study of High- and Low-Performing Students in Solving Interactive and Analytical Problems
Yiling Hu, Bian Wu and Xiaoqing Gu
300–311

Formative Assessment in Complex Problem-Solving Domains: The Emerging Role of Assessment Technologies
Kaushal Kumar Bhagat and J. Michael Spector
312–317
Middle School Students’ Mathematics Knowledge Retention: Online or Face-to-Face Environments

Clayton M. Edwards¹, Audrey C. Rule² and Robert M. Boody³

¹Grundy Center Middle School, Grundy Center, IA, USA // ²University of Northern Iowa, Cedar Falls, IA, USA // cedwards@spartanpride.net // audrey.rule@uni.edu // robert.boody@uni.edu
³Corresponding author

(Submitted March 26, 2016; Revised July 14, 2016; Accepted August 13, 2016)

ABSTRACT

Educators seek to develop students’ mathematical knowledge retention to increase student efficacy in follow-on classwork, improvement of test scores, attainment of standards, and preparation for careers. Interactive visuals, feedback during problem solving, and incorporation of higher-order thinking skills are known to increase retention, but a comparison of online and face-to-face learning regarding knowledge retention has not been fully explored. The current study tested 38 Caucasian eighth grade students for knowledge retention on ten mathematical topics they had learned in sixth grade during either online or face-to-face conditions. After two years, students were given the same posttest as in 6th grade over the 10 units of which five had been learned online and five face-to-face. Scores for long-term gain scores showed no significant differences between online or face-to-face learning conditions. A TOST calculation was used to show equivalence of middle school student long-term learning across online and face-to-face conditions in mathematics in this study within the limits of the test. The interactive nature of each condition contributed to strong mathematical understanding, which led to the retention being equivalent for both conditions.

Keywords
Mathematics, Knowledge retention, Middle school, Math, Comparison of online and face-to-face learning

Introduction

Retention of knowledge is vital for future application of concepts. Because knowledge deteriorates after long periods of non-use, it is important to determine the best ways to instruct students so that they retain information and skills (Hall, Stiles, & Horwitz, 1998). Several studies on knowledge retention have been conducted in the fields of medicine and military science in particular. In the medical profession, a patient’s life may be in the hands of a surgeon who has to make a quick decision when an unexpected complication occurs. A medical study focusing on retention of knowledge of diabetes showed that knowledge loss can happen in as little as a week if the new learning is not reinforced often (Bell, Harless, Higa, Bjork, Bjork, Bazargan, & Mangione, 2008). Another medical study (Naidr, Adla, Janda, Feberová, Kasal & Hladíková, 2004) showed that one year later, students retained only 66.8% of their previously tested learning.

In the military field, a country’s safety may hinge on a critical choice made by a soldier or group under the pressure of combat. This situation was examined in a military retention study (Ricci, Salas, & Cannon-Bowers, 1996). In this investigation, even a 25 day-lapse after learning concepts related to aviation antisubmarine warfare or basic electricity and electronics resulted in a 24% loss in knowledge of these areas. While K-12 education topics do not tend to address immediate life or death matters, knowledge retention is important for students to score well and gain college admittance to pursue their desired careers as well as being crucial to the survival of school districts in this era of high stakes testing of student knowledge.

Knowledge retention in mathematics education

Researchers in mathematics knowledge retention have attempted to identify key learning components that promote long-term learning. One investigation compared student diagram construction versus logical argument in learning geometry proofs (Lovett & Anderson, 1994), finding that work with diagrams rather than logical explanations of the problems resulted in stronger memory for rules of geometry proofs. Another study (Butcher & Aleven, 2008) examined long-term learning in which tenth grade students used a computer application that provided hints and feedback as students solved geometry problems compared to student interaction with labeled geometry diagrams. These researchers found that visual information needed to be integrated into problem-solving practice to support knowledge retention. A third investigation examined the effects of reform mathematics teaching versus more traditional mathematics instruction (Edwards & Townsend, 2012), concluding
that work involving higher-level thinking and student interaction in the problem solving process produced higher standardized test scores, which are valued data in today’s educational landscape. A comparison that has not been researched as thoroughly is long-term learning in an online environment versus a face-to-face setting, especially for middle school students. Because of this need for empirical results related to online learning and mathematical knowledge retention coupled with the increasing amounts of online learning occurring in today’s educational system (Aharoni & Bronstein, 2014; Bejerano, 2008; Young, Birtolo & McElman, 2009), the researchers conducted a study to test equivalence of the conditions. As Greener (2013) asked, is digital learning “supporting learning better than its physical precursor?” (p. 417).

Literature review

Online and face-to-face retention comparisons

Few studies have been conducted to show knowledge retention comparisons between online and face-to-face conditions, especially in the area of mathematics. The following review of the literature highlights studies comparing an online condition to a face-to-face condition regarding knowledge retention.

One study that compared knowledge retention under online versus face-to-face learning involved university medical students in an anatomy and physiology class (Rondon, Sassi, & de Andrade, 2013). In this study, students were randomly selected to attend either a course utilizing computer games as instruction (online learning), or a course involving traditional classroom (face-to-face) lecture and discussion. To gather retention data, the researchers gave a pretest before the course, an identical test at the conclusion of the nine-week course (posttest), and an identical test six-months after the conclusion of the course (long-term posttest). Regarding short-term gains from pretest to posttest, both conditions showed no significant difference in gain scores. However, the face-to-face group’s gain scores from pretest to long-term posttest revealed a significantly smaller loss of previous learning in the long-run.

Researchers of another retention study (Ricci et al., 1996) tested 60 students in a naval training course learning about chemical warfare. The students were split into three groups. One group read a set of printed lecture notes, another group studied a book, and the final group used an online computer program. Identical pretest, posttest, and retention tests (administered four weeks after the posttest) were given. Students in each condition improved in knowledge from the pretest to the posttest in a statistically significant manner ($p < .001$); however, only the online group showed a statistically significant positive difference in retention from posttest to retention test ($p < .05$). Therefore, this study, in contrast to the previously-discussed anatomy and physiology class experiment, retention favored the online condition.

In a K-12 education example, a study was conducted with a Turkish high school biology class to compare the use of online hypermedia materials with traditional lecture (Yildirim, Ozden, & Aksu, 2001). Both conditions allowed students to engage for 15 hours with the information. The students were given a pretest to set a baseline. A posttest was given immediately at the conclusion of the biology course, and the retention measurement component was administered one month after the conclusion of the class. On the long-term retention test, the control lecture group lost points compared to their previous posttest score, while the hypermedia class gained points. The difference in retention of both conditions was found to be statistically significant ($p < .05$), favoring the online condition.

A final retention study example involved a medical-surgical nursing class separated into two conditions: students instructed through a computer-assisted learning module versus those attending a more conventional lecture and demonstration (Fernandez Aleman, Carrillo de Gea, & Rodriguez Mondejar, 2011). Investigators administered a baseline pretest to determine initial knowledge. After a ten-week lapse, positive retention occurred in both conditions with no statistically significant differences in student scores.

Overall, in the current literature, there are studies that favor online or face-to-face learning or neither condition. To provide more empirical information about long-term knowledge retention regarding online mathematical learning, the authors conducted the current study. This investigation examined knowledge retention of eighth grade mathematics students who had previously experienced both online and face-to-face instruction in learning ten mathematics topics during the sixth grade.
Interactive learning and knowledge retention

Earlier theories of knowledge acquisition (Anderson, 1982; Fitts, 1964; Rasmussen, 1986; VanLehn, 1996) can be consolidated into a three-phase process model of learning (Kim, Ritter & Koubek, 2013). This model has these stages: (1) obtain declarative and procedural knowledge, (2) combine and solidify the acquired knowledge and (3) fine-tune the knowledge through overlearning. During the first stage, students learn facts and concepts that support ways of using those facts; in the second stage, this knowledge is consolidated into a procedure for applying the information; and finally, during the third stage, practice allows students to speed application of the procedures (Kim et al., 2013). The fine-tuning stage not only produces greater fluency of knowledge recall, but the number of distinct days that a student practices a particular mathematical skill is important to long-term knowledge retention (Wang & Beck, 2012). This result confirms a finding in the field of vocabulary (Pavlik & Anderson, 2005), that wide spacing of practice provides increasing benefit as practice accumulates resulting in less loss of knowledge later. Similarly, the author of a review of long-term retention of science knowledge (Custers, 2010) suggested the following strategies gleaned from the reviewed literature, to increase long-term retention of learning: (1) brief appropriately-spaced practice sessions; (2) avoidance of prolonged intensive study, instead interspersing of short bouts of intensive study; (3) periodic review of concepts already studied or mastered; and (4) regular quizzing and testing.

More-interactive methods of instruction have a positive effect on student mathematical knowledge retention. A study that involved college students who were learning Cantor set theory (Narli, 2011) found that students taught with an interactive, Constructivist methodology compared to traditional lecture, retained more mathematics information 14 months later than the control group. A study of 60 ninth grade students in Turkey (Guvercin, Cilavdaroglu, & Savas, 2014) showed that students who engaged in mathematical problem posing had better achievement, attitudes, and retention of knowledge 40 days after conclusion of instruction. Finally, a meta-analysis of 55 studies that had participants from kindergarten through college compared the use of concrete manipulatives to the control condition of discussion using abstract math symbols of concepts (Carbonneau, Marley, & Selig, 2013). These studies showed that overall there was a small to medium positive effect size of using manipulatives; a separate analysis of knowledge retention using manipulatives showed a large effect size.

Computer games can provide an active, multisensory learning environment; therefore their potential to increase student engagement and learning is great. The individualized mass and distributed practice of computerized games supports the different stages of learning in Kim, Ritter and Koubek’s theory (2013) previously described. A study of games use, compared to traditional lecture, in teaching anatomy and physiology of the head and neck to college students (Rondon et al., 2013) showed overall no significant difference between conditions for the short term, but an advantage for face-to-face instruction in the long term. Those investigators surmised that the ability of students using texts in the face-to-face condition to go back and review information as desired influenced the results and suggested that games be altered to provide this ability. In contrast, a study of fourth graders’ mathematical learning of multiplication and division of natural numbers and fractions (Pili & Aksu, 2013) showed that the experimental group using the computer software performed better than the traditionally taught control group on all areas of the posttest, and performed better in the long term on operations involving natural numbers, but not fractions. Other experimental–control group research addressing seventh grade students’ misconceptions about probability (Gürbüz & Birgin, 2012) indicated that computer-assisted teaching was significantly more effective than traditional methods on a posttest, but the researchers did not provide long-term learning effects. This attention to misconceptions can assist students in solidifying and fine-tuning their knowledge.

The current study investigates long-term knowledge retention of middle school students who participated two years earlier during sixth grade in a counterbalanced, interactive online and interactive face-to-face mathematics study in which ten mathematics topics which were part of the district’s curricular plan (prior to full implementation of the Iowa Core) were taught with students experiencing five topics under each condition. The online condition incorporated virtual manipulatives, online lecture videos, online simulations and interactive diagrams, online games and peer interactions through an online chat. Similarly, the face-to-face condition incorporated hands-on manipulatives, live lectures and explanations, visuals and diagrams, games, and group work with discussion. Both conditions were designed and taught by the same instructor. Posttest results at the end of sixth grade indicated that learning results of the two conditions were equivalent (Edwards, Rule, & Boody, 2013).
Methods

The current study examined the retention of mathematical knowledge two years after a series of posttests given in a prior study (Edwards et al., 2013) comparing the long-term knowledge retention of sixth grade (now eighth grade) students in face-to-face and online conditions in a mathematics classroom.

Participants

The participants in this study included 38 Caucasian eighth graders, 20 male, and 18 female from the same rural Iowa school district. The original sixth grade study had 46 Caucasian students; an attrition of five males and three females from the original to the current study. No additional students were added to the study. This study was approved by the committee for research on human subjects at the researcher’s university. Fully-informed, written school district, parent, and student consents to participate in the study were obtained.

Study design

The eighth graders had previously been divided into mixed-ability groups in sixth grade by the district’s fifth grade teachers with no knowledge of an eventual study. While the students (as eighth graders) were no longer grouped this way for courses, the data were analyzed using the previous groupings from sixth grade. As eighth graders, 16 of the students had their former sixth grade mathematics teacher as the primary mathematics instructor (7 male/9 female) while the other 22 students had an alternate instructor for either regular eighth grade math or Algebra I (13 male/9 female). Because this research followed a quasi-experimental design, conclusions were established on the variables implemented (Clark & Creswell, 2010); in this study, the independent variable was the type of instruction: online or traditional face-to-face teaching during sixth grade. The dependent variables were student academic retention shown by long gain scores (the change from the original sixth-grade posttest taken two years ago to the final eighth-grade posttest score) and the final eighth-grade posttest scores. The sixth grade mathematics topics were not specifically retaught during seventh and eighth grades, as the state-required standards addressed new concepts rather than revisiting the old ones.

Posttesting occurred during the last month of the students’ eighth grade year. Some sample questions taken from the probability unit include $P($multiple of 3) given a set of cards numbers 1-10 and $P(7)$ given a standard die (Foresman & Wesley, 1998, p. 628). The ten final eighth-grade posttests were the exact same tests as the students answered throughout their sixth grade year as part of the previous study (Edwards et al., 2013), and covered the topics of decimals, statistics, algebra, probability, measurement, symmetry, geometry, polygons, perimeter, and area (all topics listed in the order they were taught and decided upon by the district’s curricular focus at the time the students were in 6th grade). During the previous study, the students participated in ten units, alternating between an online condition and a face-to-face condition in a counterbalanced design. The online condition had students sit separately at desks in the classroom with individual laptop computers during mathematics class time. Learning modules were set up online that included videos and/or electronic manipulatives and interactive sites to explain concepts. Students were not allowed to communicate face to face; instead, they used a chat program to interact with the instructor and other students. Instant messaging has been shown to be an effective way to support informal mathematics coaching (Hrastini, Edman, Andersson, Kawnine, & Soames, 2014). Students completed the same book assignments and used feedback videos to check and see how well they understood the material. Students turned these assignments in electronically, digitally asked any questions they might have had, and were free to move on to the next section in the module.

Students in the face-to-face condition also sat individually at desks in the classroom. The instructor explained the concepts and utilized manipulatives and interactive activities to reinforce understanding. The teacher assigned the same book assignments as used for the online condition. Students could collaborate with other students or ask the instructor questions about the concepts verbally. Students received feedback on completed assignments the following day. Students who completed the day’s work had other enrichment or problem solving opportunities but could not move on to the next section.

The same instructor in the same classroom oversaw both the online and face-to-face environments. Two different sections of students were involved in the original study. One class learned a mathematical topic face-to-face for a few weeks, while the other class learned the same topic in an online manner. Classes switched conditions for each topic, thereby learning five of the ten topics under each condition. No homework was given for either of the conditions, allowing the focus to be on working during each class period. When online manipulatives or
interactive games were used, a similar version was presented in both conditions. For example, the online condition utilized the National Library of Virtual Manipulatives (Utah State University, 1999) to show equal fractions through the use of virtual pie pieces while the face-to-face students had physical colored pie slices to put together manually to make the comparison.

The instructional focus of both conditions was mathematics for understanding. A premium was placed on students being able to explain the underpinnings of the material being presented in both formats. To ensure the effectiveness of this approach, both conditions regularly utilized manipulatives, interactive activities, peer discussion, teacher explanations, and the requirement of students to demonstrate understanding for completed work. While the instruction occurred on video for the online condition and in person for the face-to-face environment, both situations had the same experienced instructor who explained the meaning behind the mathematics rather than memorizable procedures. This interactive pedagogical approach is necessary for deep learning to occur (Narli, 2011). Information pertaining to materials and methods used for both conditions from a portion of the probability unit can be found in Table 1.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Face-to-face condition</th>
<th>Reference</th>
<th>Online condition</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction</td>
<td>Standard Direct Instruction (Similar Content to the Video for the Online Condition) by the Teacher with Discussion</td>
<td>NA</td>
<td>Online Instruction Video Created by the Teacher</td>
<td>Edwards, 2010a</td>
</tr>
<tr>
<td>Manipulatives</td>
<td>Physical Coins and Decks of Cards</td>
<td>NA</td>
<td>Coin and Card Simulations</td>
<td>Random.org, 1998a; Random.org, 1998b</td>
</tr>
<tr>
<td>Checking the Problems</td>
<td>Teacher Walks Around Checking/Asking Questions to Students/Groups/Adding Insight When Needed</td>
<td>NA</td>
<td>Online Instruction Video Created by the Teacher for Students to Check Work and Gain Additional Understanding</td>
<td>Edwards, 2010b</td>
</tr>
<tr>
<td>Communication</td>
<td>Teacher to Student and Student to Student Conversations</td>
<td>NA</td>
<td>Teacher to Student and Student to Student to Student to Student Google Talk</td>
<td>Google, 2011</td>
</tr>
</tbody>
</table>

Original posttests were taken at the end of each trimester to compare the conditions (online and face-to-face) to see if one was better for a student’s learning than the other. The current study is an extension of the previous study, focusing on long-term retention of mathematical content and procedures learned under each condition.

**Instrumentation**

The ten posttests each contained ten questions and were adapted from the *Middle School Math Course One* sixth grade textbook used in the school district (Foresman & Wesley, 1998). This text followed the state and local standards. The posttests each contained ten questions of varying difficulty levels. Students needed to demonstrate mathematical understanding for a response to be scored as correct. Tests were untimed, on paper, and scored by the teacher, as in the original study to ensure scoring continuity.

**Data analysis**

Student long gain scores and new, final eighth-grade posttest scores were analyzed topic by topic. The scores of the class experiencing the traditional face-to-face method were compared to scores of the class learning through the online methods for the same topic. A t-test was applied to scores earned under both learning conditions for each of the mathematics topics to determine if there were significant differences in learning between long gain scores and new posttest scores.
In the previous study, students participated in five of the ten topics for each learning situation (online or face-to-face). The five scores from each condition were totaled to generate a composite score for each student. Each composite score had a maximum of 50 possible points (a 10 on each of the five sections of the pretest-posttest results in a total score of 50). Using this information, the equivalence of retention for online learning of mathematics and retention of face-to-face learning of mathematics was tested.

Results and discussion

Long gain scores by mathematics topic

The long gain scores (the change from the original sixth-grade posttest taken two years ago to the final eighth-grade posttest score) for all ten units of instruction are shown in Table 2. These long gain scores are fairly small because during sixth grade, most students had mastered much of the mathematics material presented and did not have as much room for growth as they originally did. Table 2 shows three places at which students evidenced negative gain scores (a loss of previously-learned content). Students learning perimeter online and the topics of area and polygons under the face-to-face condition forgot some of the mathematical concepts they had previously successfully applied during the sixth-grade posttests. The perimeter concept students had the most difficulty remembering concerned the idea that a circle’s perimeter is approximately three times the diameter. Students who originally studied area in the face-to-face condition seem to have forgotten some of the material; they especially scored low on complex area problems of calculating area of figures with straight and circular sides. The students who originally studied these concepts in the online condition may have had an advantage in working with more interactive online diagrams that allowed students to disassemble the complex figure and analyze each part.

There were no statistically significant differences between long gain scores of eighth grade students who originally studied these topics during the online and face-to-face conditions of the original sixth-grade study. This indicates that, in general, students taught in an online condition or in a face-to-face condition seemed to have built a similarly strong foundation of knowledge in these topics and to have retained similar amounts of mathematical information.

Table 2. Long gain scores for each separate topic

<table>
<thead>
<tr>
<th>Order</th>
<th>Mathematics topic</th>
<th>Trimester taught</th>
<th>Class group</th>
<th>n</th>
<th>Mean long gain score</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Online</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>Decimals</td>
<td>1</td>
<td>A</td>
<td>19</td>
<td>1.11</td>
<td></td>
<td>.48</td>
<td>.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>.79</td>
<td></td>
<td>2.35</td>
<td>.03</td>
</tr>
<tr>
<td>2nd</td>
<td>Statistics</td>
<td>1</td>
<td>A</td>
<td>19</td>
<td>2.18</td>
<td></td>
<td>.71</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>1.55</td>
<td></td>
<td>2.19</td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>Probability</td>
<td>1</td>
<td>A</td>
<td>19</td>
<td>.53</td>
<td></td>
<td>2.22</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>1.50</td>
<td></td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>Algebra</td>
<td>1</td>
<td>A</td>
<td>19</td>
<td>1.76</td>
<td></td>
<td>1.55</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>1.00</td>
<td></td>
<td>2.80</td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>Measurement</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>.47</td>
<td></td>
<td>1.70</td>
<td>.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>1.00</td>
<td></td>
<td>1.83</td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td>Symmetry</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>.21</td>
<td></td>
<td>1.40</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>.21</td>
<td></td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>7th</td>
<td>Geometry</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>.74</td>
<td></td>
<td>1.33</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>1.16</td>
<td></td>
<td>2.12</td>
<td></td>
</tr>
<tr>
<td>8th</td>
<td>Polygons</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>.11</td>
<td></td>
<td>2.00</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>.79</td>
<td></td>
<td>2.15</td>
<td></td>
</tr>
<tr>
<td>9th</td>
<td>Perimeter</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>-.53</td>
<td></td>
<td>1.61</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>.47</td>
<td></td>
<td>2.93</td>
<td>.51</td>
</tr>
<tr>
<td>10th</td>
<td>Area</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>-.47</td>
<td></td>
<td>2.93</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>.26</td>
<td></td>
<td>2.33</td>
<td></td>
</tr>
</tbody>
</table>

Final eighth-grade posttest scores by mathematics topic

Table 3 lists the new final eighth-grade test scores for each of the ten different topics. The only unit of instruction that showed a statistically significant difference between the two conditions regarding final eighth-grade scores
was the unit on probability. Class A (originally in the online condition for this topic in sixth grade) scored quite a bit lower than Class B (a difference of 1.31 points on the 10-point test). Both classes scored very similarly on probability on the original sixth-grade posttest (Class A’s score was 5.79; Class B’s score was 6.13).

<table>
<thead>
<tr>
<th>Order</th>
<th>Mathematics topic</th>
<th>Trimester taught</th>
<th>Group</th>
<th>n</th>
<th>Mean long gain score</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>Decimals</td>
<td>1</td>
<td>A</td>
<td>19</td>
<td>7.63</td>
<td>1.67</td>
<td>-1.12</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>8.21</td>
<td>1.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>Statistics</td>
<td>1</td>
<td>A</td>
<td>19</td>
<td>6.53</td>
<td>2.37</td>
<td>-0.30</td>
<td>.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>6.74</td>
<td>1.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>Probability</td>
<td>1</td>
<td>A</td>
<td>19</td>
<td>6.32</td>
<td>1.86</td>
<td>-2.18</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>7.63</td>
<td>1.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>Algebra</td>
<td>1</td>
<td>A</td>
<td>19</td>
<td>8.05</td>
<td>2.04</td>
<td>-0.60</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>8.42</td>
<td>1.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>Measurement</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>6.89</td>
<td>1.63</td>
<td>-0.82</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>7.37</td>
<td>1.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td>Symmetry</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>8.11</td>
<td>2.08</td>
<td>.08</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>8.16</td>
<td>1.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th</td>
<td>Geometry</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>6.74</td>
<td>2.10</td>
<td>.07</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>6.83</td>
<td>2.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th</td>
<td>Polygons</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>7.37</td>
<td>1.61</td>
<td>1.64</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>6.26</td>
<td>2.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9th</td>
<td>Perimeter</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>5.84</td>
<td>2.04</td>
<td>-0.42</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>6.11</td>
<td>1.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10th</td>
<td>Area</td>
<td>3</td>
<td>A</td>
<td>19</td>
<td>3.63</td>
<td>1.83</td>
<td>-1.31</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>19</td>
<td>4.42</td>
<td>1.90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparing the total of all online to the total of all face-to-face scores by student

In this part, long gain score totals will be discussed first, followed by final eighth-grade posttest score totals. The reader will recall that the posttest consisted of a set of ten tests, one on each mathematics topic. In the original study, students learned five topics online and five topics face-to-face. Because the study was counterbalanced, topics learned online by one class were learned face-to-face by the other class. A paired-samples t-test was conducted comparing each student’s total of the long gain scores for the appropriate five mathematical topics for the online condition with the total of the long gain scores for the appropriate five mathematical topics learned in the face-to-face control condition. It was nonsignificant (Mean of the paired differences = .11, t(37) = .10; p > .05, Standard Error of the Mean = 1.05).

Recognizing that a lack of significance does not necessarily imply equivalence but may simply indicate insufficient power (Hoenig & Heisey, 2001), the researchers went further to directly test equivalence concerning the long gain score totals. Following the TOST (two one-sided t-tests) approach (see, for example, Chen, Rathore, Ji, & Germansderfer, 2010), we first developed a “zone of indifference”: the width of scores within which scores could be taken as equivalent. This is not itself a statistical matter; it should be developed directly connected to their understandings of mathematics. Based on his report, we chose ± 2.5 points (out of the 50 points each student could have earned across 5 topics). Thus, for the score 40, 42 would be considered equivalent, but 43 would not be considered equivalent to 40.

The zone of indifference for the long gain scores, then, would be 1.5 pts. ± 2.5 pts., or -1.0 to 4.0. We then implemented the two one-sided t-tests approach by developing a 90% confidence band around the mean of the difference scores. For the gain scores, this band was .54 ± 1.50 ± 2.46 (the mean of the differences (1.50) ± .96 (which is the standard error of the mean (.78) x 90CI t(37) (1.687)). This suggests that the long gain scores for online and face-to-face were not equal (statistically different) but can be considered equivalent within the level of indifference chosen.

The zone of indifference for the gain scores, then, would be 1.5 pts. ± 2.5 pts., or -1.0 to 4.0. We then implemented the two one-sided t-tests approach by developing a 90% confidence band around the mean of the difference scores. For the gain scores, this band was .54 ± 1.50 ± 2.46 (the mean of the differences (1.50) ± .96 (which is the standard error of the mean (.78) x 90CI t(37) (1.687)). This suggests that the long gain scores for online and face-to-face were not equal (statistically different) but can be considered equivalent within the level of indifference chosen.

For the new, eighth-grade posttest score totals, a paired-samples t-test between online (total of the appropriate 5 new test unit scores) and traditional face-to-face (total of the appropriate 5 new test unit scores) was calculated. It was significant (Mean of the paired differences = 1.50, t(37) = 2.61, p = .013, Standard Error of the Mean
In this sample, at least, students who were originally taught online in sixth grade, compared to students originally learning face-to-face in sixth grade, had higher final eighth-grade test scores in the current study two years later. Although the results of this test indicated a significant difference in eighth-grade posttest scores between students who had originally learned these topics online or face-to-face during sixth grade, this difference was small and may possibly fit within the zone of indifference for the TOST. Therefore a TOST was conducted on eighth grade posttest score totals. This statistical difference may be influenced by loss of subjects: the original sixth-grade study involved 46 students, but by eighth grade, 8 of these students had moved to other schools, eliminating them from the eighth-grade posttesting.

For the TOST on the new eighth-grade posttest scores, the zone of indifference would be .11 ± 2.5 pts., or -2.39 to 2.61. The band provided by the TOST analysis was -1.66≤ 0.11≤1.88 (the mean of the differences (.11) ±1.77 (which is the standard error of the mean (1.05) x 90CI t(37) (1.687)). Because the entire confidence interval in each of these two cases (for gain scores and for posttest scores) is contained within the zone of indifference, and there was no statistical difference from this TOST, the researchers conclude that the achievement as measured by the new eighth-grade posttest scores from the online and the traditional learning conditions may be considered as equivalent at the .05 level of significance.

Conclusion

Summary

This study was conducted to determine whether retention of mathematics knowledge in an online mathematics environment was equivalent to retention of mathematics knowledge in a face-to-face mathematics environment for middle school students as shown through posttest performance two years later. Individual t-tests used on each topic comparing the online condition with the face-to-face condition in regards to long-term gains (retention from sixth grade to eighth grade) showed no significant difference for any topic.

Next, a t-test comparing student long-term gain (retention) and a t-test comparing new posttest scores under both conditions (total scores for five mathematical topics) were used showing no significant difference in long-term gain, but a significant difference in final posttest scores in favor of online learning. Two additional TOST analyses were conducted to evaluate equivalency for both long-term gain scores and for final posttest scores. The results of the TOST indicated that the difference occurred between the confidence intervals, meaning that both the online and face-to-face conditions may be considered equivalent for both long-term gain scores and final posttest scores within the established zone of indifference. This information shows that online learning can be a viable method of learning for middle school students in regards to both mathematical knowledge retention and achievement.

Interactive learning

The instruction sixth grade middle school students experienced during the face-to-face and online conditions had many interactive components including virtual or hands-on manipulatives; online simulations with interactive diagrams or visuals and diagrams with live explanations and discussion; online games or hands-on games; and peer interactions via a chat center or face-to-face group work. These interactions allowed students to satisfy the three phases of Kim et al.’s (2013) model of learning. Students learned facts and concepts from online or live lectures and textbook readings, then, they combined and solidified that knowledge through interactive manipulatives, simulations, diagrams, and problems solved with peers via group work in chats or face-to-face, and finally, they fine-tuned their knowledge through additional practice during online or face-to-face games. Both the online and face-to-face conditions supported student learning through interactivity contributing to equivalent knowledge retention across the conditions and supported the findings of Carbonneau, Marley, and Selig (2013), Guvercin, Cilavdaroglu, and Savas (2014), and Narli (2011).

Recommendations for future research

While this study showed mathematical retention in middle school students to be equivalent in both online and face-to-face conditions, more questions remain. Similar studies expanded to different student age groups might indicate optimal ages for online learning. The current study showed middle school students evidenced equivalent mathematical retention in both online and face-to-face situations; would younger populations produce similar
results? Is there a point when students are too young to enjoy the benefits that mathematical online learning can bring? Students in the original sixth grade study (Edwards et al., 2013), who had experienced both conditions, reported enjoying online learning more than face-to-face instruction as long as the online learning was interspersed with face-to-face interactions. Do students feel the same way in the long term? Does a more prolonged exposure to online learning throughout middle school and high school cause that enjoyment to wane?

Acknowledgements

This research project was supported in part by the Center for Educational Transformation at the University of Northern Iowa, Cedar Falls, Iowa.

References


Recommending Learning Activities in Social Network Using Data Mining Algorithms

Lamia Mahnane
LRS Laboratory, University of Badji Mokhtar, Annaba, Algeria // mahnane_lamia@yahoo.fr

(Submitted April 8, 2016; Revised September 27, 2016; Accepted October 19, 2016)

ABSTRACT
In this paper, we show how data mining algorithms (e.g., Apriori Algorithm (AP) and Collaborative Filtering (CF)) is useful in New Social Network (NSN-AP-CF). “NSN-AP-CF” processes the clusters based on different learning styles. Next, it analyzes the habits and the interests of the users through mining the frequent episodes by the Apriori algorithm. Finally, it groups dynamically the users based on the collaborative filtering. The participants in this study consisted of 80 university students who were asked to analyze the differences in skill level when using various learning activities. Moreover, 40 students were included in this study in order to examine the effectiveness of NSN-AP-CF. The experiment results proved that the proposed algorithm, which considers the grouping dynamically the users and the discovery of all frequent episodes, generates better precisions compared with the other algorithms ($F1 = 0.649$).

Keywords
Social networks, Data mining, Apriori algorithm, Learning style, Collaborative filtering

Introduction
Social network has become one of the comfortable medium for user to share their knowledge (Zaidieh, 2012). Users pay more attention to share their knowledge spontaneously in a relaxed, informal environment more than the formal classroom environment. A teacher can easily understand user’s learning from outside of the class and they can get full details about the learning system. It is very helpful for a teacher to understand the difficulties of the user he/she facing in the learning system.

Learning through social networking sites is limited when different users have different preferred ways to learn. Some may understand quickly through games and simulations, others may prefer problem solving. Some may deal well with theories, others may learn through projects and examples (Veletsianos & Navarrete, 2012).

In this way, most researches are based on using different characteristics to group users forgetting that these characteristics can change at any time. This change of the characteristics leads to the problem of the static grouping where the system does not edit the created groups automatically. In Mahnane and Touati (2015), the authors investigated the relationship between Traditional educational Social Networking (TSN) and learning styles.

Furthermore, information from social networks can present valuable data to report student problem. Examining such data, however, can be challenging task. The problem of student’s behaviors reveal from social network site need human analysis. There are many traditional methods available such as questionnaires, surveys and interviews to analyze the student’s behaviors in educational social networking sites (Chen, Vorvoreanu, & Krishna, 2014). But the main problem with these methods is these techniques cannot be performed efficiently with big data as the analysis has to be performed manually. For this, data mining techniques collect and analyze data generated in social networking sites.

The main objective of this research is to develop an educational social network site based on data mining techniques and learning styles. The rest of the paper is organized as follows. Section “Background and related scientific work” introduces the related work. The design of educational social network based on data mining methods and learning styles is described in Section “Research methods.” Section “Analysis and Results” is reserved to present the tests and the obtained results. The conclusion and the future works are drawn in Section “Conclusion and future prospective.”
Background and related scientific work

This section is organised in three subsections, firstly the general context of the educational social networks is briefly introduced. Secondly, we focus on the grouping of students based on learning style. We conclude with using data mining in educational social networks.

Educational social networks

The social networking sites brought about major change in how communication and participation between users and communities and information exchange (Allen, 2012; Greenhow, Gleason & Li, 2014). Advantages of using social network in education are:

- **Participation**: Social network encourages contributions and feedback from everyone who is interested.
- **Openness**: Most social networks are open to feedback and participation. They encourage voting, comments and the sharing of information.
- **Conversation**: Social network is better seen as a two way conversation.
- **Community**: Social network allows communities to form quickly and communicate effectively.
- **Communities**: Share common interests.
- **Connectedness**: Most kinds of social network thrive on their connectedness, making use of links to other sites, resources and users.

Thus, educational social network can provide the teacher and students a space in which they can discuss their experiences and their lessons. The Educational social network helps the teacher, student establish a long lasting relationship and powerful interactions with each other. These interactions help them determine educational needs. A general architecture of educational social networking inspired from (Valova, 2015), is shown in Figure 1.

![Figure 1. A general architecture of educational social networking inspired from (Valova, 2015)](image)

The following is an explanation of the general components of educational social networking sites (Mahnane & Touati, 2015).
• **Events**: reminders of appointments and activities that we must implement. When the merger will always remember these dates because they will always be with us, both when you open the educational environments, or in times of entertainment across social networking sites.

• **Chat/Messages**: In the educational environment to be trapped between the student and the teacher only or between the student and the last in the means of social communication to be with a group of individuals, whether teachers or students are generic.

• **Latest News**: A list containing the news from people and groups on social network or educational environment, latest events include.

• **Profile**: through personal files can identify the name of the person, find out basic information about it, such as race, and date of birth, and the interests and personal images in addition to other information, is a profile entry gate to the world of the person, it is through the main page of the profile can be seen activity person recently, and find out who his friends are and what are the new images placed in addition to other activities.

• **Course**: the educational environment is set up course through the introduction of the name and the article is here from the field, but most important of these fields is the material name. And then it is handled through a set of tools that contribute to the educational process, including raising the files, images, video, questions, plus students and teachers of the material, and the distribution of the roles on them.

Moreover, the algorithm integrated educational environment and a social network (Mahnane & Touati, 2015) is summarized as follows:

**Algorithm integration**

```
Begin
Take the input of educational environment and social network.
Read the educational environment components from set of components collection.
For each content educational environment do
  Begin
  Compared the educational environment components with social network components
  If components educational environment equal components social network, add content components educational environment to components social network.
  End
End.
```

Here are some examples of the social networks that used in education:

• In Rožac, Pogačnik, Kos, Buendía, and Ballester (2012), the authors proposed an integration of e-learning systems with social networks and display its supporting software. The author solved the low level of interaction between users. Through direct relationship between learning content and communication between users and teachers in e-learning systems. Suggested use of social networks to increases the interaction between users in e-learning environments. The approach depends on the virtual classroom, integrating e-learning system COOM with Facebook.

• In Du, Fu, Zhao, Liu, and Liu (2012), the authors proposed an interactive and collaborative e-learning platform which integrates social software with a learning management system (LMS). This platform provides personalized space for users where they can interact and collaborate with others. The personalized space of users contains their course network, social network and knowledge network. This platform connects course network of users with his/her social network and knowledge network. Furthermore, users are able to build their personalized social network and knowledge network during the process of learning.

• In Meishar-Tal, Kurtz, and Pieterse (2012), the authors used a Facebook as an alternative to LMS. Their approach reviews the current research on the use of Facebook in academia and analyzed the differences between a Facebook group and a regular LMS. The authors used a Facebook group as a course website, serving as a platform for delivering content and maintaining interactions among the students and between the students and the lecturer.

• In Kurtz (2014), the author study the effect of integrating Facebook group and course website on participation and perceptions on learning. Such that use of two virtual platforms for learning. Show that Facebook, can be used for discussion and exchange of knowledge. Students reported that Facebook helps enhance the interaction and social learning processes with emphasis on the involvement of the user, and contribute effectively, and frequent interaction with peers and the instructor.
Finally, the differences between previous researches and our approach are as follows:

- Developing a model to enable creation, registration course student and teacher in educational environment.
- Developing a model of social networking site.
- Developing a model to enable creation, storing, publication and sharing of course materials from educational environment to social network.
- We will apply different learning methods to get an optimal teaching and learning strategy. This strategy based on learning style and data mining techniques depends on integrating of educational environments, and social networks.

**Educational social networks and learning styles**

Not all users can be assumed to benefit from social networks either due to their diverse backgrounds or due to their different learning styles, it may not be appropriate to suggest that social networks might be beneficial for every student, as students are generally from diverse backgrounds, and most importantly have different learning styles (García-Martín, & García-Sánchez, 2013; Lin, Hou, Wang, & Chang, 2013; Friesen & Lowe, 2012).

Felder and Solomon (2001) described learning styles across four dimensions answering the following questions: What type of information is emphasized by the instructor (Sensing (SE)/Intuiting (IN))? What mode of presentation is stressed (Visual (VI)/Verbal (VE))? What mode of student participation is facilitated by the presentation (Active (AC)/Reflective (RE))? What type of perspective is provided on the information presented (SeQuential (SQ)/Global (GL)).

In this way, Mahnane, and Touati (2015) used learning styles in order to provide interactive social network delivery by means of adaptable navigation and adaptive content selection. The most important contributions in this work are as follows: (1) Designing Traditional Social Network (TSN) based on user learning style and user skill level; (2) Supporting different types of educational content. In Mahane and Touati’s (2015) study, the authors have shown that the use of TSN gives a good result compared to Facebook. Figure 2 shows an interface for TSN based on user learning style (VI/SQ/IN/AC) and user skill level (average).

![Figure 2. An integrated in TSN: domain (Mathematical logic), learning style (VI/SQ/IN/AC) and skill level (average)](image)

**Educational social networks and data mining**

Nowadays social networks provide an important source of information. In addition to the common use, they are also used by researchers to extract information that is usually not visible to the naked eye. Analyzing such data, however, can be challenging. Within the most important educational social network from the fields of data mining we can find:

- In Gamila, Pavla, Jan, Katerina and Václav (2010), the authors were presented an application of spectral clustering method to find the patterns of behavior of groups of students enrolled in the e-learning system. For easier generation of the graphs described students’ behavioral patterns in e-learning system, with requires setting of number of input variables for clustering method or setting of the dimensions (selection of the appropriate course or students’ activities in the course provided in e-learning system), was developed specialized software. Moreover, authors attempted to find relations between the behavioral study patterns...
and the students’ study performance in the selected course. The findings of the experiment did not show any relation between the similarity in students’ behavior and their grades as well as relation between students’ positions in generated network and their academic performance.

- In Ambrósio Gomes, Cavalcante Prudêncio, Azevedo Filho, Alves do Nascimento, and Alves de Oliveira (2013), the authors presented the evaluation of their strategy to group profiling applied to an educational social network called OJE. The OJE is a social network that connects students and teachers through games and enigmas. The OJE’s platform provides an environment for conducting tournaments between teams and individual students (supervised by teachers) who engage in various disputes. In order to explain the group formation resulted from OJE educational social network, the authors present two differentiation-based group profiling methods: Wilcoxon rank-sum test and PART rules Algorithm.

- In Troussas, Espnosa, and Virvou (2016), the authors described the affect recognition for intelligent language learning using Rocchio Classifier. Furthermore, the authors presented important features for achieving a probabilistic approach of Rocchio classifier. The significance of using a more probabilistic approach of Rocchio algorithm for affect recognition is that the probabilistic methods are preferable from a theoretical viewpoint, since a probabilistic framework allows the clear statement and easier understanding of the simplifying assumptions made. The used data is a random sample of streaming Facebook states and were not collected by using specific queries. The size of hand-labeled data allows performing cross validation experiments and checking for the variance in performance of the classifier across folds. In this way, knowing the emotional state of each user, the authors used this characteristic as a value of the vector used for the group profiling, which can further ameliorate the educational experience through Facebook.

Discussion

From the Background and related scientific work, the following problems should be recognized:

- The majority of the traditional educational social network do not consider the adaptation according to the user’s profile in their approaches (Troussas, Espnosa, & Virvou, 2016; Ambrósio Gomes et al., 2013; Gamila, Pavla, Jan, Katerina & Václav, 2010).
- Less attention was intended to study the application of machine learning techniques for generation of group profiling in educational social networks (Ambrósio Gomes et al., 2013).
- Several researches are made in the group formation where all these researches are vocalized on grouping users using different criteria: users’ profiles, learning styles, etc. These previous works present some limitations where the grouping of users is static while the characteristics of users are dynamic (Mahnane & Touati, 2015). As result, groups are formed with the initial and the previous characteristics and not with the current ones.

A summary of the related scientific research is presented in Table 1.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Social network</th>
<th>Communities detection</th>
<th>Static/dynamic</th>
<th>Identification of learning activities</th>
<th>Recommendation process</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ambrósio Gomes et al., 2013)</td>
<td>OJE</td>
<td>Explicit</td>
<td>Static</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>(Troussas, Espnosa, &amp; Virvou, 2016)</td>
<td>Facebook</td>
<td>Explicit</td>
<td>Static</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>(Mahnane &amp; Touati, 2015)</td>
<td>TSN</td>
<td>Implicit</td>
<td>Static</td>
<td>Pedagogical rules</td>
<td>Pedagogical rules</td>
</tr>
<tr>
<td>(Rožac et al., 2012)</td>
<td>Facebook</td>
<td>Explicit</td>
<td>Static</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>(Meishar-Tal et al., 2012)</td>
<td>Facebook</td>
<td>Explicit</td>
<td>Static</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>(Kurtz, 2014)</td>
<td>Facebook</td>
<td>Explicit</td>
<td>Dynamic</td>
<td>Apriori algorithm</td>
<td>Collaborative Filtering</td>
</tr>
<tr>
<td>Proposed approach</td>
<td>NSN-AP-AF</td>
<td>Implicit</td>
<td>Static</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 1. A summary of the related scientific research
Research methods

System framework

The operating principle of our model NSN-AP-CF as well as the input and output of each module is shown in Figure 3 and Table 2 respectively.

Furthermore, we describe in the following the different steps of our approach:

Step 1: The system gathers the users in different groups based on the learning styles of users.

Step 2: A pre-test questionnaire was given to the students for evaluates the users’ skill level about the mathematical logic course.

Step 3: The users can access to their own lesson where they can access to the learning activities. Every lesson contains many activities such as: theory (A1), games and simulations (A2), problem solving (A3), discussion (A4), case study (A5), question/answer method (A6), project (A7) and practical work (A8).

The system selects a frequent episodes by the Apriori algorithm.

Step 4: The system evaluates the user’s skill level for each lesson

Step 5: The system updates groups of users based on collaborative filtering algorithm.

![Figure 3. Framework for NSN-AP-CF using apriori algorithm and collaborative filtering](image)

<table>
<thead>
<tr>
<th>Module</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning style module</td>
<td>Questionnaire, Index Learning Style</td>
<td>Style in four dimensions</td>
</tr>
<tr>
<td>Mining sequential patterns module</td>
<td>Style in four dimensions, Log file</td>
<td>The most probable activities</td>
</tr>
<tr>
<td>Evaluation module</td>
<td>Questions, Log file</td>
<td>Users’ ratings</td>
</tr>
<tr>
<td>Recommendation process module</td>
<td>Log file, users’ ratings file</td>
<td>16 clusters. Learning style based on skill level for user</td>
</tr>
</tbody>
</table>

The detail of each module in the proposed model NSN-AP-CF is presented as following.

Learning style identification

In our research, we employ the instrument inspired by Felder-Silverman’s ILS to examine user’s learning style by an online questionnaire, with the paraphrases on items. By literature review on learning styles and the
 experimental conclusion about Felder-Silverman’s model (Felder & Solomon, 2001; Mahnane & Hafidi, 2016),
the following general activities are provided for each pole of the 4 dimensions of our instruments (see Table 3).

Table 3. Learning styles and corresponding learning activities inspired from (Mahnane & Hafidi, 2016)

<table>
<thead>
<tr>
<th>Learning style</th>
<th>Learning activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active (AC)</td>
<td>• Games and simulations</td>
</tr>
<tr>
<td></td>
<td>• Problem solving</td>
</tr>
<tr>
<td></td>
<td>• Discussion</td>
</tr>
<tr>
<td></td>
<td>• Project</td>
</tr>
<tr>
<td>Reflective (RE)</td>
<td>• Case study</td>
</tr>
<tr>
<td></td>
<td>• Question/answer</td>
</tr>
<tr>
<td>Sensing (SE)</td>
<td>• Problem solving</td>
</tr>
<tr>
<td></td>
<td>• Question/answer</td>
</tr>
<tr>
<td>Intuitive (IN)</td>
<td>• Games and simulations</td>
</tr>
<tr>
<td></td>
<td>• Discussion</td>
</tr>
<tr>
<td></td>
<td>• Case study</td>
</tr>
<tr>
<td></td>
<td>• Project</td>
</tr>
<tr>
<td>Visual (VI)</td>
<td>• Games and simulations</td>
</tr>
<tr>
<td>Verbal (VE)</td>
<td>• Discussion</td>
</tr>
<tr>
<td></td>
<td>• Question/answer</td>
</tr>
<tr>
<td>Sequential (SQ)</td>
<td>• Question/answer</td>
</tr>
<tr>
<td>Global (GL)</td>
<td>• Case study</td>
</tr>
<tr>
<td></td>
<td>• Project</td>
</tr>
</tbody>
</table>

Mining by Apriori algorithm

The mining by apriori algorithm establishes connection between episodes learning and the choice of activity. With this relationship, we can detect users’ preference, as shown in Table 4.

Table 4. A small part of the database

<table>
<thead>
<tr>
<th>User-id</th>
<th>Users’ learning style</th>
<th>Users’ rating</th>
<th>Users’ gender</th>
<th>Units</th>
<th>Lessons</th>
<th>Access-time</th>
<th>Episode learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VI/GL/AC/IN</td>
<td>2</td>
<td>Female</td>
<td>1</td>
<td>1</td>
<td>2015. Jan 1</td>
<td>{A2, A5, A1, A4}</td>
</tr>
<tr>
<td>1</td>
<td>VI/GL/AC/SE</td>
<td>3</td>
<td>Male</td>
<td>1</td>
<td>1</td>
<td>2015. Jan 1</td>
<td>{A2, A3, A8, A1}</td>
</tr>
<tr>
<td>1</td>
<td>VE/GL/AC/IN</td>
<td>4</td>
<td>Female</td>
<td>1</td>
<td>1</td>
<td>2015. Jan 12</td>
<td>{A7, A5, A1, A6}</td>
</tr>
<tr>
<td>2</td>
<td>VE/GL/RE/IN</td>
<td>2</td>
<td>Female</td>
<td>1</td>
<td>1</td>
<td>2015. Jan 20</td>
<td>{A5, A1, A3}</td>
</tr>
<tr>
<td>2</td>
<td>VE/SQ/RE/SE</td>
<td>3</td>
<td>Male</td>
<td>1</td>
<td>1</td>
<td>2015. Jan 21</td>
<td>{A1, A6, A8, A3}</td>
</tr>
<tr>
<td>2</td>
<td>VE/GL/RE/IN</td>
<td>3</td>
<td>Male</td>
<td>1</td>
<td>1</td>
<td>2015. Jan 22</td>
<td>{A5, A1, A3}</td>
</tr>
<tr>
<td>3</td>
<td>VI/GL/AC/IN</td>
<td>2</td>
<td>Female</td>
<td>1</td>
<td>1</td>
<td>2015. Jan 15</td>
<td>{A2, A4, A5, A1}</td>
</tr>
<tr>
<td>3</td>
<td>VI/GL/RE/IN</td>
<td>4</td>
<td>Female</td>
<td>1</td>
<td>1</td>
<td>2015. Jan 15</td>
<td>{A5, A1, A4, A3}</td>
</tr>
</tbody>
</table>

Table 5. The results of mining by Apriori algorithm (best rules)

<table>
<thead>
<tr>
<th>Learning styles</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI/GL/AC/IN</td>
<td>{A3, A2, A1}</td>
<td>{A7, A5, A1}</td>
</tr>
<tr>
<td>VI/GL/AC/SE</td>
<td>{A8, A2, A1}</td>
<td>{A8, A5, A1}</td>
</tr>
<tr>
<td>VI/GL/RE/IN</td>
<td>{A2, A1, A3}</td>
<td>{A5, A1, A7}</td>
</tr>
<tr>
<td>VI/GL/RE/SE</td>
<td>{A2, A1, A3}</td>
<td>{A5, A1, A8}</td>
</tr>
<tr>
<td>VI/SQ/AC/IN</td>
<td>{A3, A2, A1, A4}</td>
<td>{A3, A5, A6, A1}</td>
</tr>
<tr>
<td>VI/SQ/AC/SE</td>
<td>{A3, A2, A4, A1}</td>
<td>{A8, A5, A4, A1}</td>
</tr>
<tr>
<td>VI/SQ/RE/IN</td>
<td>{A2, A1, A3}</td>
<td>{A5, A1, A7}</td>
</tr>
<tr>
<td>VI/SQ/RE/SE</td>
<td>{A2, A4, A1, A3}</td>
<td>{A5, A4, A1, A8}</td>
</tr>
<tr>
<td>VE/GL/AC/IN</td>
<td>{A3, A2, A1}</td>
<td>{A7, A5, A1}</td>
</tr>
<tr>
<td>VE/GL/AC/SE</td>
<td>{A3, A2, A4, A1}</td>
<td>{A8, A5, A4, A1}</td>
</tr>
<tr>
<td>VE/GL/RE/IN</td>
<td>{A2, A4, A1, A3}</td>
<td>{A5, A4, A1, A7}</td>
</tr>
<tr>
<td>VE/GL/RE/SE</td>
<td>{A2, A4, A1, A3}</td>
<td>{A5, A4, A1, A8}</td>
</tr>
<tr>
<td>VE/SQ/AC/IN</td>
<td>{A7, A2, A1}</td>
<td>{A8, A5, A1}</td>
</tr>
<tr>
<td>VE/SQ/AC/SE</td>
<td>{A3, A2, A1}</td>
<td>{A7, A5, A1}</td>
</tr>
<tr>
<td>VE/SQ/RE/IN</td>
<td>{A2, A1, A3}</td>
<td>{A5, A1, A7}</td>
</tr>
</tbody>
</table>
Afterward, with using this algorithm, we can predict the most probable episode learning and activities. The mining by Apriori algorithm is defined as follows:

**Algorithm: Apriori Algorithm;**

**Begin**

\( C_1 = \text{Itemsets of size one in I}; \)

Determine all large itemsets of size 1, \( L_1; \)

\( i=1; \)

Repeat

\( i=i+1; \)

\( C_i = \text{Apriori-Gen}(L_{i-1}); \)

Apriori-Gen\( (L_{i-1}) \)

Generate candidates of size \( i+1 \) from large itemsets of size \( i. \)

Join large itemsets of size \( i \) if they agree on \( i-1. \)

Prune candidates who have subsets that not large

Count \( C_i \) to determine \( L_i; \)

Until no more large itemsets found;

**End**

In Table 5, we show the results of mining sequential patterns by Apriori algorithm with min-supp (\( \alpha = 0.4 \)) and min-conf (\( \delta = 0.9 \)).

**Dynamic grouping of users**

The objective of this work consists of forming groups of users with different learning styles. For this, we define the k-mean algorithm based on users learning style and users' skill level as follows:

**Algorithm K-mean;**

**Input** Maximum number of clusters \( k; \)

Activity (unit, lesson, difficulty level, learning style, type of activity, time);

User (goal and preferences, learning style, skill level, gender);

**Begin**

1. Choose \( k \) individuals randomly (as a center of initial classes);

2. Assign each individual to the nearest center; This gives a partition \( P \) into \( k \) classes \( P1=\{C1, C2,..., Ck\}; \)

3. The centers of gravity are calculated for each class \( P1 \), which provides \( k \) new cluster centers.

4. Repeating step (2) and (3) until two successive iterations yield the same partition.

**End.**

In Table 6, we show the skill level had a significant relationship in the group G12. From this result, the users’ learning styles (VI, RE, GL, and SE) had a better skill level in the episode {games and simulations, problem solving}.

**Table 6. The centers of gravity in 16 clusters**

<table>
<thead>
<tr>
<th>Centers of gravity</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
<th>G7</th>
<th>G8</th>
<th>G9</th>
<th>G10</th>
<th>G11</th>
<th>G12</th>
<th>G13</th>
<th>G14</th>
<th>G15</th>
<th>G16</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>VE</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>SQ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IN</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Several empirical findings have suggested that the sensitive style had a better skill level in the discussion (Huang, Lin, & Huang, 2012), and the global style had a high skill level in the learning based on problem solving (Carmo, Gomes, Pereira, & Mendes, 2006).
Methodology

In this experiment, we wished to ascertain if the use of an educational social network (NSN) was an effective way to improve student skill level and course satisfaction in “mathematical logic” course when compared to traditional social network (TSN).

Design of the study

The student skill level was measured by the pre/post-test instrument. Both the experimental and the control groups experienced similar learning activities. However, only the experimental group received access to the new educational social network (NSN). The performance pre-test served to determine the degree of homogeneity between the experimental and control groups. During the end of session the same performance post-test was administered to all participants. Additional data were also collected, in the form of post surveys (Brooke, 1996). The experimental groups post-survey explored the efficacy and the overall usefulness of NSN for learning.

Participants

To validate our approach, an experiment was conducted at the department of mathematic, university of Algeria, where the students learn the concepts of “Mathematical logic” subject. These students were divided into experimental and control groups. Each group consisted of 40 students. The NSN system was designated for the experimental group and the TSN system was designated for the control group. Figure 4 shows an interface for problem solving activity in new educational social network (NSN) based on Apriori algorithm and collaborative filtering (K-mean algorithm).

Figure 4. Problem solving activity in NSN-AP-CF

Analysis and results

Students’ skill level

The first research question investigates whether differences between the experimental and control groups exists in terms of the students’ skill level in the “mathematical logic” course between the students who used the NSN versus TSN, as measured by pre and post instrument. Table 7 presents the descriptive statistics for the pre-test and post-test data for both the experimental and control groups. A review of the descriptive statistics reveals that the pre-test scores appeared to be similar across groups; however, the post-test scores appeared to be different across groups with students in the experimental group outperforming students in the control group.

A t-test analysis was performed for both the pre-test and post-test scores function of the group type to determine the existence of differences between the two groups’ means. For the pre-tests, the analysis concluded that there is not a statistically significant difference between the means of the pre-test scores of the control group and the experimental group (p < .05).
On the other hand, for the post-test scores, the analysis revealed that there is a statistically significant difference between the mean post-test scores for the two groups considered.

Based on the descriptive statistics generated in Table 7 and the conclusion of the t-test we can determine if indeed the utilization of the NSN had a positive impact over the post-test scores – the students in the experimental groups scored higher than the ones in the control group. The results of the hypothesis testing conducted states that indeed the mean of the post-test scores for the students included in the experimental group is statistically higher than the one of the students from the control group ($p < .05$). So, we can conclude that the NSN increased the students’ skill level in their final session.

<table>
<thead>
<tr>
<th>Table 7. Descriptive statistics for pre-test and post-test measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Male ($n = 14$)</td>
</tr>
<tr>
<td>Female ($n = 26$)</td>
</tr>
</tbody>
</table>

We also calculate the difference between the post-test scores and pre-test scores, to determine if the students from the experimental group performed better than the ones assigned to the control group. The $t$-test concluded that indeed the students from the experimental group performed better than the ones in the control group ($p < .05$).

To support these findings, that indeed the students exposed to the NSN performed better than the ones not using the system, we conducted statistical testing on the learning activities for the two groups. Even though at the beginning of the course there was no statistical difference between the activities scores for the experimental versus the control group, by the last testing we were able to identify that the our approach is efficient. The analysis results are synthesized in Table 8.

<table>
<thead>
<tr>
<th>Table 8. A comparison between the experimental and control groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activities</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Total score</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unit1/lesson1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unit1/lesson2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unit2/lesson1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unit2/lesson2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unit3/lesson1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unit3/lesson2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unit3/lesson3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unit3/lesson4</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

These results have been supported by:

- Kunchala (2015) thought that student’s posts on social network gives us a better concern to take decision about the particular education system’s learning process of the system.
- Buzzetto-More (2012) Concluded that social networking have been shown to foster social learning while engaging students in a complex array of communicative and creative endeavors including new literacy practices.
- Shaqifah, Sani, Taib, Jusoff, and Shazi (2011) alleged that data mining discovered the patterns of students’ participation in social networking. It is found that their participation relates with their personal behavior.
Students’ perception

The second research question addresses the experimental groups’ level of perceived satisfaction regarding the effectiveness of the NSN tools after the experiment; to what extent do students perceive access to the NSN contributed to their learning. The experimental groups’ responses to the post-survey items referring to the usefulness of the NSN overall were used to answer this research question. The data for these 10 survey items were coded as 5 - Strongly Agree or 1 - Strongly Disagree. Descriptive statistics for the survey items 1-10 indicated that the perception scores appeared to be high, as displayed in Table 9. Table 9 shows that Q10, Q5 and Q1 provide the best score (Q10 = 64%, Q5 = 59.2%, Q1 = 70.7%).

<table>
<thead>
<tr>
<th>Table 9. Descriptive statistics for experimental - Survey response items</th>
<th>Mean</th>
<th>SD</th>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>I think that I would like to use this system frequently</td>
<td>3.94</td>
<td>1.028</td>
<td>5.3</td>
</tr>
<tr>
<td>Q2</td>
<td>I found the system unnecessarily complex</td>
<td>3.56</td>
<td>1.132</td>
<td>7.6</td>
</tr>
<tr>
<td>Q3</td>
<td>I needed to learn a lot of things before I could get going with this system</td>
<td>3.18</td>
<td>1.066</td>
<td>12.0</td>
</tr>
<tr>
<td>Q4</td>
<td>I felt very confident using the system</td>
<td>3.49</td>
<td>1.050</td>
<td>5.3</td>
</tr>
<tr>
<td>Q5</td>
<td>I would imagine that most people would learn to use this system very quickly</td>
<td>3.65</td>
<td>1.019</td>
<td>5.3</td>
</tr>
<tr>
<td>Q6</td>
<td>I thought there was too much inconsistency in this system</td>
<td>3.52</td>
<td>0.835</td>
<td>3.1</td>
</tr>
<tr>
<td>Q7</td>
<td>I found the various functions in this system were well integrated</td>
<td>3.67</td>
<td>0.972</td>
<td>3.1</td>
</tr>
<tr>
<td>Q8</td>
<td>I found the system very cumbersome to use</td>
<td>3.41</td>
<td>1.011</td>
<td>5.3</td>
</tr>
<tr>
<td>Q9</td>
<td>I think that I would need the support of a technical person to be able to use this system</td>
<td>3.65</td>
<td>1.029</td>
<td>3.1</td>
</tr>
<tr>
<td>Q10</td>
<td>I thought the system was easy to use</td>
<td>3.85</td>
<td>0.915</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Note. * [score of “Strongly Agree” and “Agree” is low]

Precision methods

The third research question addresses the precision of our proposed approach by a comparative study. Table 10 and Figure 5 presents a comparative study between the APriori algorithm (AP), Collaborative Filtering algorithm (CF), Traditional Social Network based on learning styles system (TSN), and APriori and Collaborative Filtering algorithm (AP-CF). For this, we can define the notions of precision, recall and F-measure (Aher & Lobo, 2012) as follow:

\[ F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \]

\[ \text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \]

\[ \text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \]

As can be seen, the AP-CF algorithm that considers the grouping dynamically the users based on their learning styles and the discovery of all frequent episodes based on Apriori algorithm generates better precisions compared with the other algorithms.

<table>
<thead>
<tr>
<th>Table 10. The comparison between TSN, NSN-AP, NSN-CF, and NSN-AP-CF algorithms</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSN</td>
<td>0.297</td>
<td>0.598</td>
<td>0.396</td>
</tr>
<tr>
<td>NSN-AP</td>
<td>0.399</td>
<td>0.624</td>
<td>0.486</td>
</tr>
<tr>
<td>NSN-CF</td>
<td>0.497</td>
<td>0.646</td>
<td>0.561</td>
</tr>
<tr>
<td>NSN-AP-CF</td>
<td>0.566</td>
<td>0.758</td>
<td>0.649</td>
</tr>
</tbody>
</table>
Conclusion and future prospective

Some researchers have shown that the users instead of educational environments and the social networks are more effective than users of educational environments (Veletsianos, & Navarrete, 2012; Troussas, Espnosa, & Virvou, 2016). Indeed, the social networks facilitate the involvement of the user process and make it in the heart of the educational process. Furthermore, social networks use several components that interact with the user more than others. Thus, the integration of educational environments with the social networks will have the ability to improve, support and build privileged system.

This paper describes a new educational social network based on Apriori and Collaborative Filtering algorithms (NSN-AP-CF) which can automatically adapt to the interests, learning styles and knowledge levels of users. Moreover, it aims at grouping dynamically the users based on user’s learning style and user’s skill level.

The experiment results confirm that a mining the frequent episodes by the Apriori algorithm in each learning style has the potential to improve the quality of NSN-AP-CF, as well as keep the recommendation up-to-date. Moreover, this study shows that our proposed approach can outperform the traditional recommendation algorithms significantly in precision and could be more suitable for personalization of educational environments. Different users can benefit from this study as follows:

- Motivate users by providing new and modern means of learning and teaching.
- Make it easy for users to access the information required in the fastest time.
- Flexibility and ease, students communicate with teachers and with peers material with enriches the learning process.

As a future work, we plan to conduct other tests in order to validate the proposed approach with a large sample of users. In addition, we envisage extracting the benefits of our approach by applying it to other domains. Finally, users can still use social networking tools for other purposes rather than learning.

References


Kurtz, G. (2014). Integrating a Facebook group and a Course Website: The Effect on participation and perceptions on learning. American Journal of Distance Education, 28(4), 253-263. doi:10.1080/08993647.2014.957952


Veletianaos, G., & Navarrete, C. C. (2012). Online social networks as formal learning environments: Learner experiences and activities. The International Review of Research in Open and Distance Learning, 13(1), 144-166.

Analyzing Pauses in Computer-Assisted EFL Writing—A Computer-Keystroke-Log Perspective

Cuiqin Xu* and Yan Qi
Nanjing University, Nanjing, China // cxq@nju.edu.cn // yzqiyan@sina.com
*Corresponding author

(Submitted April 10, 2016; Revised July 5, 2016; Accepted August 15, 2016)

ABSTRACT
Using computer keystroke logs, this study investigated how writing skill affected L2 writers’ pausing patterns to gain insights into their management of the cognitive writing processes. The 59 participants, 29 in the more-skilled group and 30 in the less-skilled group, were recruited from a college English course at a key Chinese university. The two groups completed an argumentative essay in a computer classroom where Inputlog6.0 was installed to log their writing activities. Setting the pause threshold at 2 seconds, the study examined both the global pausing patterns and the interval pausing patterns by dividing each writing event into five equal intervals, and how the final text quality related to the pausing patterns. The results showed a significant effect of writing skill on the interval pausing patterns, but not on the global pausing patterns. Correlating significantly with the final text quality, the interval pausing patterns also revealed important differences in L2 writers’ management of writing processes in terms of how one writing process dominates at specific intervals with interaction and shifts between other processes in a recursive fashion. Pedagogical implications are then discussed in light of these findings.

Keywords
L2 writing skill, Pauses, Writing process, Writing intervals

Introduction
Written text production involves complex cognitive processes that place heavy demands on the writers’ working memory (Kellogg, 1996; Olive, Kellogg, & Piolat, 2008). When writers experience cognitive overload in their working memory, some processes would be suspended or even sacrificed to accommodate the immediate call for a specific process (DeKeyser, 2001; de Larios, Manchón, & Murphy, 2006) and pausing is a strategy that writers actively or passively adopt to free up attentional resources for processes of immediate priority. Therefore, pauses may serve as windows to writers’ writing activities, such as “phrasing, memory such, decision, feedback, conceptual integration, and so forth” (de Beaugrande, 1984, p. 166). In addition, although writing consists of major processes as planning, translating (i.e., putting ideas into visible language) and revising (Hayes & Flower, 1980), these processes are not activated linearly but interact with one another in a recursive fashion, with one process dominating while other processes remaining dormant until being reactivated. While Olive and Kellogg (2002) suggest skilled writers can activate transcription concurrently with translating, planning and revising are mainly activated during pauses (Alves, Castro, & Olive, 2008; Olive, Alves, & Castro, 2009). Therefore, pausing may reveal possible problems and writing strategies behind such shifts of writing processes (Wengelin, 2006). As writing performance depends on how writing processes are organized during the composition process (Van den Bergh & Rijlaarsdam, 2007), pausing, being windows to writing activities and cognitive processes during text production, merits more research attention if the writing process is to be better understood.

Many pause studies were conducted in pen-and-paper settings and mostly adopted video-recording or think-aloud protocols to probe into the writing process. While both methods have produced enlightening results for the comprehension of writing and think-aloud protocols have, in particular, advanced the knowledge of how pauses are filled qualitatively, these methods are limited to case studies due to practical reasons. In addition, there have been wide concerns over the intimidating nature of video-recording and the reactivity issues of thinking-aloud in writing research (for a review, see Yang, Hu, & Zhang, 2014).

With computer arising to be the leading writing medium for academic purpose, writing has witnessed much difference from that in pen-and-paper settings (Alves, Castro, Sousa, & Strömqvist, 2007; Olive & Kellogg, 2002). Meanwhile, with the advent of computer science and technology, a number of keystroke logging tools have been developed and greatly promoted written production research (Leijten & Van Waes, 2006). These computer programs “log and time stamp keystroke activities to reconstruct and describe the text production processes” (Leijten & Van Waes, 2013, p. 359), thus enabling writing research to examine the online writing process both multi-dimensionally and at higher accuracy levels. More importantly, these programs, by running in the background, minimize writers’ anxiety of being observed. Informed by recent development in writing
research and theory, these programs also provide rather sophisticated analyses of the writing process, thus more opportunities for refined writing research (Latif, 2009; Sullivan & Lindgren, 2006; Van Waes, Leijten, Wengelin, & Lindgren, 2012).

As different writing processes draw on the same working memory pool (Kellogg, 2001; McCutchen, 2000), the activation or deactivation of certain processes provides insights into the writers’ management of their cognitive processes during the text production. With the rationale behind keystroke logging that “writing fluency and flow reveal traces of the underlying cognitive processes” (Leijten & Van Waes, 2013, p. 360), this study, using Inputlog6.0 to log L2 writing activities, endeavors to examine how pauses provide insights into L2 writers’ management of their writing process in computer-assisted writing settings.

**Literature review**

**Pauses in pen-and-paper writing settings**

Early writing research identifies three major processes in writing, i.e., planning, translating and revising (Hayes & Flower, 1980) and suggests that the ability to consciously manage these writing processes is a fundamental component of writing skill, contributing much to producing good-quality texts (Levy & Ransdell, 1995). Many studies gained insights into these cognitive processes by probing into pauses during text production. Early pause studies in pen-and-paper writing settings mainly used video-taping or think-aloud protocols and had long pause thresholds for practical reasons. Schumacher, Klare, Cronin and Moses (1984), setting the pausing threshold at 10 seconds, videotaped and compared the pausing of 22 high school graduates and 20 college students through a 30-minute writing assignment. The study found no differences in pause frequency between the two groups, but the high school group paused averagely longer than the college group. Dividing the writing events into four intervals, the researchers reported that two groups did not differ significantly in pause duration at different writing intervals. However, more cognitive activities were documented for the college group while more grammatical activities were observed for the high school group during the pauses.

Many pausological studies have focused on the relationship between pause duration and the grammatical locations of such pauses, i.e., within-word, between words/clauses/sentences or at T-unit boundaries (Matsuhashi, 1987; Spelman Miller, 2000; Wengelin, 2006). These studies generally suggest that pause duration increases with the grammatical unit level, i.e., pauses within a word are shorter than those proceeding a word, and pauses between sentences are shorter than those between paragraphs. Other researchers reported similar findings that grammatical, discourse and morphological boundaries affect pause length (Nottbusch, Grimm, Weingarten, & Will, 2005; Spelman Miller, 2006). Some pausological studies tried to relate writing pauses with writing processes during the composition process. Using think-aloud protocols and reaction-time tasks, Beauvais, Olive and Passerault (2011) examined pauses at different writing stages and reported that text quality positively correlated with pause length at the prewriting stage. However, as there are wide concerns over to what extent think-aloud protocols would affect the writing process (see Yang, Hu, & Zhang, 2014) and whether the dual-task mode (reaction-time task and composition task) would genuinely reflect the actual writing process, further research into the writing processes through pauses are in need.

**Pauses in computer-assisted writing settings**

With the development of computer keystroke logging tools, pausological studies have been able to log and observe the writing process in a non-intruding manner and therefore have greatly pushed the research boundaries imposed by research instruments. Van Waes and Schellens (2003) compared the pausing behavior of experienced writers in computer keyboard settings against pen-and-paper settings. Setting the pausing threshold at 3 seconds and dividing writing into segments of 10 minutes, the study found that long pauses evenly distributed throughout the writing event in computer settings while half of the long pauses concentrated at the beginning stage of writing in paper settings. The researchers thus concluded that writers tend to begin writing sooner in keyboard settings than in paper settings and that the use of a word processor would result in more fragmented writing processes. As differences have been clearly documented in pausing patterns in pen-and-paper writings against computer keyboard writings, more studies are in need to better understand the nature of writing with computer keyboard.

Mixed research results have also been reported with regard to how writing skill affects pause frequency and pause duration in computer-assisted writings. Spelman Miller (2000) reported that L2 writers paused longer in
all grammatical locations compared with L1 writers, suggesting that pause duration is a function of both language proficiency and writing skill. Spelman Miller, Lindgren, and Sullivan (2008) further examined pause frequency and pause duration of L2 writers through a 3.5-year longitudinal study. The study found that as their writing abilities increased, L2 writers paused less frequently, but their pause duration did not change significantly over time, suggesting that writing ability affected pause frequency but not pause duration. Furthermore, Xu and Ding (2014) reported that while skilled and less-skilled L2 writers did not differ in their pausing patterns from a global perspective, skilled L2 writers paused less frequently and significantly longer at the prewriting stage than their less-skilled peers, showing that writing skill affected pause frequency and pause duration at different functional stages of writing. The study also found that text quality significantly correlated with the prewriting pause duration, rendering it important to examine pause in a more refined manner than from a global perspective.

To sum up, while previous pause studies adopt various and relatively long pause thresholds due to practical reasons, computer logging tools enable L2 writing researchers to examine pauses of shorter thresholds for better understandings of L2 writers’ pausing behavior as well as their process management behind these pauses. In addition, although direct comparisons of pausing patterns at different stages or time segments are insightful, it is undeniable that a writer’s pausing pattern carries much personal traits and the management of writing process is a function of personal knowledge system and writing expertise. Therefore, it would be insightful to examine how L2 writers’ pausing patterns vary over the temporal development of writing, as indicated by the pausing patterns at different time intervals of the writing event. Such a probe could neutralize the effect of writer difference and gain more understanding of the L2 writing processes. However, precise estimate of pausing from an inter-writer perspective is still lacking.

This study attempts to investigate how pauses vary in terms of frequency and duration at different time intervals to gain insights into how writers temporally manage their writing processes. Specifically, this study attempts to investigate whether writing skill affects the pausing pattern globally and at different intervals of the writing event to gain insights into the cognitive processes underlying the written text production. For these ends, this study attempted to address the following research questions:

- Does writing skill affect EFL writers’ global pausing patterns?
- Does writing skill affect EFL writers’ pausing patterns at different intervals?
- How do EFL writers’ global and interval pausing patterns relate to the quality of their final texts?

**Method**

**Participants**

The participants of the study (*N* = 59) were recruited from a pool of 118 sophomores who enrolled in a general college English course at a Chinese University. The students, between 18 to 20 years old, had been learning English for an average of 10 years, ranging from 8 to 11 years. Roughly, the students were of intermediate and upper-intermediate levels, or B1 and B2 as described in *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Council of Europe, 2011). For the English course in which data were collected for this study, the 118 students were assigned into four classes according to the alphabetical order of their surnames. They regularly had four hours of English classes per week, two hours of listening and speaking plus two hours of reading and writing. As a routine, the participants had their writing sessions in a computer classroom, where each had a desktop computer at their disposal. The participants were fairly proficient with typewriting on computer as they had compulsory computer classes since primary schools with a focus on operation skills. In addition, college students in China, with few exceptions, have laptop computers and routinely hand in e-copies for their course work.

Upon entering college, all students took the English placement test, which included four multiple-choice sections of 85 points in total and a writing section of 15 points. Among the 118 students, those who ranked the top 35 in the writing section formed the more-skilled group and those who ranked the bottom 35 formed the less-skilled group. Eventually, 59 out of the 70 students who completed the study entered data analysis, with 29 in the more-skilled group and 30 in the less-skilled group. The 59 students participated in the study in their original classes during data collection and were only regrouped and combined for data analyses.

An independent-samples *t*-test was run on the 59 participants’ writing scores in the placement test and the results suggested that the more-skilled group (*M* = 11.26, *SD* = 1.04) and the less-skilled group (*M* = 6.32, *SD* = .90) differed significantly in their writing scores (*t* = 19.22, *p* = .00 < .05, Cohen’s *d* = 5.08).
Procedures

All data were collected in a computer classroom, where the students regularly had their reading and writing classes. Before the course began, a writing log program, Inputlog6.0, was installed on each computer and for each writing session, the students were asked to start the program and input the necessary personal data, i.e., name, age, gender, group, years of English learning. Before the data collection for this study, the students had completed two writing tasks using Inputlog and become familiar with writing in Inputlog.

At the beginning of the class when data were collected for this study, one drunk driving case reported by a local newspaper was briefly discussed in class and then the essential elements for argumentation learned earlier in their reading class were reviewed. Then, the students were asked to develop an argumentative essay on *whether drunk drivers should be imprisoned on their first offense*. They were told to develop a complete piece of argumentation on the given topic, with no requirement about the specific length of the final product. The students were asked to finish their writing in class, within a maximum length of 90 minutes.

With a click on the “Start Recording” button, a normal word processor popped up and the students began their writings, with Inputlog6.0 logging their writing events meanwhile. After finishing their writing, the students could simply click “Stop Recording” and the program would automatically exit and save the word documents in a pre-specified file folder. The instructor then collected the student writings through a management software pre-installed on all computers.

After the data collection, all students were informed of the research purpose of this study and signed the agreement of participation on a voluntary basis. As a return, one of the researchers offered the students individualized face-to-face discussion of their writings.

Data coding and analysis

*Global analysis of pauses*

Following previous pausological studies (Wengelin, 2006; Xu & Ding, 2014), this study also set the pause threshold at 2 seconds. Each writing event was analyzed globally in terms of event time, active writing time, pausing time, global pause frequency and pause duration. As the writing event time was different for each participant, raw global pause frequency was converted into standardized global pause frequency of 60 minutes through the following formula:

\[
\text{Standardized global pause frequency} = \frac{\text{Raw global pause frequency}}{\text{event time (seconds)}} \times 3600 \text{ (seconds)}
\]

*Interval analysis of pauses*

Each writing event was divided into five temporal intervals of equal length by setting “Number of Intervals” at 5 in Inputlog6.0. Pause frequency and pause duration were compared respectively at each interval between the two groups and within each group across the five intervals. As the interval length varied from one writer to another, raw interval pause frequency was converted into standardized interval pause frequency of 10 minutes by following the formula below:

\[
\text{Standardized interval pause frequency} = \frac{\text{Raw interval pause frequency}}{\text{Interval length (seconds)}} \times 600 \text{ (seconds)}
\]

*Text scoring*

Before scoring, all texts were coded with consecutive numbers to remove student identities. The writing rating rubric (see Appendix) was adapted from CET4 (College English Test, Band 4), a nation-wide English test in China. Two experienced college EFL teachers, who did not teach any of the students before or during data collection and had extensive experience in CET4 writing scoring, conducted the scoring separately. As the maximum score for a text is 15, the text would be returned to the two raters for rescoring if the score discrepancy exceeded three points.
The inter-rater reliability was checked using Pearson correlation coefficients. The initial inter-rater reliability was $r = .83$ ($p = .00 < .05$) and it reached $r = .90$ ($p = .00 < .05$) after the rescoring for score discrepancy over three points. Then, the scores from two raters were averaged as the final text score for further analysis.

**Statistical analysis**

A series of independent-samples $t$-test were performed for possible between-group difference both globally and at different intervals, and repeated measure ANOVAs were conducted for the within-subject differences across the five intervals. In addition, Pearson correlation coefficients were computed to examine how writing pauses related to the final text quality. For all statistical tests, the significance level was set at .05.

**Results**

**Global pausing patterns**

Table 1 presents an overview of time allocation and pausing patterns for the two groups. The independent-samples $t$-test results suggest that the two groups did not differ significantly with regard to the total event time ($t = -1.64, p = .11 > .05$), active writing time ($t = -1.17, p = .25 > .05$), total pausing time ($t = -.68, p = .50 > .05$), global pause frequency ($t = -.22, p = .83 > .05$) or global pause duration ($t = -.06, p = .95 > .05$), suggesting that the two groups did not differ significantly in terms of time allocation or pausing patterns from a global perspective.

<table>
<thead>
<tr>
<th></th>
<th>Less-skilled ($n = 30$)</th>
<th>More-skilled ($n = 29$)</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total event time</td>
<td>3808.10</td>
<td>4061.45</td>
<td>-1.64</td>
<td>.11</td>
</tr>
<tr>
<td>Active writing</td>
<td>2053.77</td>
<td>2195.62</td>
<td>-1.17</td>
<td>.25</td>
</tr>
<tr>
<td>Total pausing</td>
<td>1754.33</td>
<td>1865.83</td>
<td>-.68</td>
<td>.50</td>
</tr>
<tr>
<td>Global pause freq</td>
<td>213.10</td>
<td>215.17</td>
<td>-.22</td>
<td>.83</td>
</tr>
<tr>
<td>Global pause dur</td>
<td>7.74</td>
<td>7.78</td>
<td>-0.06</td>
<td>.95</td>
</tr>
</tbody>
</table>

**Interval pausing patterns**

**Interval pause frequency**

Table 2 summarizes the statistics of standardized pause frequency at the five intervals respectively. The independent-samples $t$-test results suggest that the less-skilled group paused significantly less frequently than the more-skilled group at Interval 2 ($t = -2.34, p = .02 < .05$) and the two groups were not significantly different in pause frequency in any other intervals.

<table>
<thead>
<tr>
<th></th>
<th>Less-skilled ($n = 30$)</th>
<th>More-skilled ($n = 29$)</th>
<th>$t$</th>
<th>$p$</th>
<th>Cohen’s $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval 1</td>
<td>34.69</td>
<td>30.61</td>
<td>1.82</td>
<td>.07</td>
<td>--</td>
</tr>
<tr>
<td>Interval 2</td>
<td>33.32</td>
<td>38.60</td>
<td>-2.34</td>
<td>.02*</td>
<td>.61</td>
</tr>
<tr>
<td>Interval 3</td>
<td>36.42</td>
<td>36.48</td>
<td>0.03</td>
<td>.98</td>
<td>--</td>
</tr>
<tr>
<td>Interval 4</td>
<td>38.77</td>
<td>38.74</td>
<td>0.01</td>
<td>.99</td>
<td>--</td>
</tr>
<tr>
<td>Interval 5</td>
<td>34.43</td>
<td>34.91</td>
<td>0.21</td>
<td>.83</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. *$p < .05$.

For how each group differed in pause frequency across the five intervals, repeated measures ANOVAs were conducted with the within-subjects factor being time intervals and the dependent variable being the standardized interval pause frequency. Figure 1 plots the pause frequency at each writing interval for the two groups. For the less-skilled group, Mauchly’s test indicated that the assumption of Sphericity was not violated, $X^2(9) = 10.23, p > .05$. Tests of within-subjects effects show that there was significant effect of time interval on pause frequency, $F(4, 116) = 2.71, p = .03 < .05$. $\eta^2_p = .09$, suggesting that 9% of the variation in pause frequency of...
the less-skilled group was accounted for by time interval. Pairwise comparisons suggest that the less-skilled group paused significantly more frequently at Interval 4 than at Interval 1, Interval 2 and Interval 5.

Figure 1. Pause frequency at the five intervals (Mean ± S.E.)

As for the more-skilled group, Mauchly’s test indicated that the assumption of Sphericity was violated, \(X^2(9) = 18.52, p < .05\); therefore, degrees of freedom were corrected using Huynh-Feldt estimates of Sphericity (\(\varepsilon=.87\)). Tests of within-subjects effects show that there was significant effect of time interval on pause frequency, \(F(3.47, 97.03) = 9.47, p = .00 < .05\), \(\eta^2_p = .25\), suggesting that 25% of the variation in pause frequency of the more-skilled group was accounted for by time interval. Pairwise comparisons suggest that the more-skilled group paused significantly less frequently at Interval 1 than in other four intervals and paused significantly more frequently at Interval 2 than at Interval 4.

Interval pause duration

Table 3 presents the statistics for pause duration at the five intervals. The independent-samples \(t\)-test show that the more-skilled group paused significantly longer than the less-skilled group at Interval 1 (\(t = -2.52, p = .02 < .05\)), but significantly shorter than the less-skilled group at Interval 2 (\(t = 2.17, p = .04 < .05\)).

Table 3. Interval pause duration

<table>
<thead>
<tr>
<th>Interval</th>
<th>Less-skilled (n = 30)</th>
<th>More-skilled (n = 29)</th>
<th>(t)</th>
<th>(p)</th>
<th>Cohen’s (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval 1</td>
<td>7.34</td>
<td>2.68</td>
<td>10.11</td>
<td>5.38</td>
<td>-2.52</td>
</tr>
<tr>
<td>Interval 2</td>
<td>9.28</td>
<td>3.93</td>
<td>7.43</td>
<td>2.45</td>
<td>2.17</td>
</tr>
<tr>
<td>Interval 3</td>
<td>7.70</td>
<td>2.26</td>
<td>7.60</td>
<td>2.83</td>
<td>0.14</td>
</tr>
<tr>
<td>Interval 4</td>
<td>7.45</td>
<td>2.83</td>
<td>7.45</td>
<td>3.05</td>
<td>0.00</td>
</tr>
<tr>
<td>Interval 5</td>
<td>8.16</td>
<td>4.19</td>
<td>8.04</td>
<td>5.27</td>
<td>0.10</td>
</tr>
</tbody>
</table>

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

For how each group differed in pause duration across the five intervals, repeated measures ANOVAs were conducted with the within-subjects factor being time interval and the dependent variable being pause duration. Figure 2 plots the pause duration at each writing interval for the two groups.

Figure 2. Pause duration at the five intervals (Mean ± S.E.)
For the less-skilled group, Mauchly’s test indicated that the assumption of Sphericity was violated, $X^2(9) = 28.64, p < .05$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of Sphericity ($\varepsilon = .61$). Tests of within-subjects effects show that there was no significant effect of interval on pause duration, $F(2.44, 70.66) = 2.69, p = .06, \eta^2 = .09$, suggesting that the effect of interval on pause duration approached the significance level. As Howell (2010) suggests that most of the multiple comparison procedures do not require an overall significant ANOVA group effect (pp. 372-373), pairwise comparisons were conducted. The results suggest that pause duration at Interval 2 were significantly longer than those at Interval 1, Interval 3 and Interval 4, with no significant difference in pause duration between any other two intervals.

For the more-skilled group, Mauchly’s test indicated that the assumption of Sphericity was violated, $X^2(9) = 58.28, p < .05$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of Sphericity ($\varepsilon = .59$). Tests of within-subjects effects show that there was significant effect of interval on pause duration, $F(2.37, 66.44) = 3.76, p = .02 < .05, \eta^2 = .12$, suggesting that 12% of the variation in pause duration of the more-skilled group was accounted for by time interval. Pairwise comparisons suggest that pause duration at Interval 1 were significantly longer those at Interval 2, Interval 3 and Interval 4, but not longer than that at Interval 5. This shows that Interval 1 is significantly different from the middle three intervals, displaying possible differences of the more-skilled group in managing their writing processes at different intervals.

In summary, the pausing patterns of more-skilled and less-skilled writers mainly differed during the first two intervals. The more-skilled group paused less frequently and longer at Interval 1, but paused shorter and more frequently as the writing went on; by contrast, the less-skilled group displayed more frequent yet shorter pauses at Interval 1, and less frequent yet longer pauses at Interval 2. In addition, both groups witnessed a slight increase in pause duration and a decrease in pause frequency at Interval 5.

### The final text quality and pausing patterns

Table 4 presents the writing product data for the two groups in terms of process words (number of words produced during the process), product words (number of words in the final product) and the final text quality (text score). As shown in Table 4, compared with the less-skilled group, the more-skilled group wrote significantly more words during the process ($t = -2.84, p = .01 < .05$), produced significantly longer final texts ($t = -.46, p = .00 < .05$) and their final texts were of significantly better quality ($t = -7.49, p = .00 < .05$).

<table>
<thead>
<tr>
<th>Table 4. Writing product data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Less-skilled (n = 30)</strong></td>
</tr>
<tr>
<td><strong>More-skilled (n = 29)</strong></td>
</tr>
<tr>
<td><strong>M</strong></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Process words</td>
</tr>
<tr>
<td>Product words</td>
</tr>
<tr>
<td>Text score</td>
</tr>
</tbody>
</table>

*Note.* $p < .05$.

Two-tailed Pearson correlation coefficients were computed to see how the global and interval pausing patterns related to the final text quality. The results suggest that the final text quality did not correlate significantly with either global pause frequency ($r = -.08, p = .57 > .05$) or global pause duration ($r = .13, p = .33 > .05$). However, the final text quality correlated positively with pause duration at the first interval ($r = .38, p = .00 < .05$) but negatively with pause frequency at the first interval ($r = -.25, p = .05$). No significant correlation was observed between the final text quality and pause frequency or pause duration at other intervals.

To sum up, the more-skilled group were significantly more productive than the less-skilled group during the writing event and produced significantly better final texts. Significant correlations were observed between the final text quality and pauses at the first interval, showing that the final text benefited from long and infrequent pauses at the beginning of the writing event.

### Discussion

This study set out to examine how writing skill affected writing pauses to gain insight into L2 writers’ management of their writing processes. The answers to the research questions have specified important pausing patterns both globally and at different temporal intervals and how pausing patterns related to the final text quality.
First, more-skilled and less-skilled L2 writers did not differ significantly with regard to writing time allocation or global pausing patterns, showing that writing skill does not affect the intermediate L2 writers’ allocation of writing time from a global perspective. This confirms Xu and Ding’s (2014) finding that L2 writers of intermediate level did not distinguish themselves in the global pausing patterns, testifying to some commonality in their management of L2 writing processes. Together with the fact that the final text quality did not correlate with either global pause frequency or global pause duration, this study suggests the global pausing pattern is not a good indicator of either writing skill or final text quality. However, the fact that the two groups differed significantly in the final text length and quality suggests more productive use of the writing time for the more-skilled group, lending support to Ferrari, Bouffard, and Rainville’s (1998) finding that poor college students wrote shorter texts than their peers. Pedagogically, this study shows writing productivity is not a simple function of writing time and L2 writers need instruction on more qualitative use of their writing time from a global perspective.

Second, more-skilled and less-skilled L2 writers differed in their interval pausing patterns and the final text quality significantly correlated with pauses during the first interval, showing that writing skill affects L2 writers’ management of writing processes and consequently, their writing products. Spelman Miller (2000) reported that pause duration increased at larger grammatical unit boundaries and signaled significant planning (p. 142). In this study, the interval pausing patterns suggest the more-skilled group may engage in much global planning at Interval 1, displaying long and infrequent planning pauses; at Interval 2, this group engaged in more focused translating of ideas, displaying short and more frequent translating pauses. By contrast, the less-skilled group displayed frequent and short pauses in Interval 1, suggesting that this group started writing much sooner and Interval 1 related to both planning and translating. Moving to Interval 2, the less-skilled group displayed less frequent but longer pauses, showing that their focused translating process was frequently interrupted by other processes. While Spelman Miller et al.’s (2008) found that global pause frequency declined but global pause duration did not change as writing abilities developed, this study documented differences in the interval pausing patterns between more-skilled and less-skilled L2 writers and therefore suggests the importance of examining pauses at fine-grained temporal intervals.

The positive correlation between pause duration at Interval 1 and the final text quality supports Beauvais et al.’s (2011) finding on pen-and-paper writings that text quality benefits from long pauses at the prewriting stage. In addition, Ferrari et al. (1998) reported that long episodes of pausing during translating emerges as a result of insufficient planning at the prewriting stage. When writing breaks down, writers would have to strive for global conceptualization and planning in the middle of translating. As writing is constantly restructured and redefined by the produced text, insufficient global planning could impose extra cognitive load on working memory with assessing and integrating new content into the existing text. Since text quality depends on the ability to plan globally before actual drafting and use this plan to guide the writing process (Chai, 2006; De La Paz & Graham, 2002; MacArthur, Harris, & Graham, 1994; Xu & Ding, 2014), L2 writers should be pedagogically reassured about the importance of global planning at the beginning of writing and that a hasty start of drafting does not gain them advantage in their final products.

Thirdly, this study documents significant effects of time interval on pausing patterns, manifesting transitions of writing processes in a recursive fashion along the temporal development of writing. The pauses of the more-skilled group at Interval 1 were significantly different from other intervals except the last one. Informed by early writing models (Hayes & Flower, 1980; Kellogg, 1996), this pausing pattern clearly marks out the planning, translating and revising stages during the composition process. By contrast, the less-skilled writers paused significantly longer at Interval 2 than other intervals except the last one. This suggests that the less-skilled group had much shorter initial planning process and began their translating process much sooner (Ferrari et al., 1998). When their writing encountered breakdowns, the less-skilled group had to engage in further planning at Interval 2, followed by focused translating at Interval 3 and Interval 4 and revising at Interval 5. Therefore, the less-skilled group experienced planning, translating, planning, translating and revising in their composition process.

Although the writing process is highly interactive and recursive among various component processes (Hayes & Flower, 1980; Kellogg, 1996), this study suggests an effect of writing skill on how different writing processes dominate and shift at different intervals: the more-skilled group competently managed their writing processes with well-defined functional stages, while the less-skilled group exhibited less competent management of writing processes and juggled back and forth with poorly-defined functional stages. This confirms previous findings from think-aloud protocols that more-skilled writers concentrated more on formulation (translating in Hayes & Flower’s model) during the central stages of writing (de Larios et al., 2006; de Larios, Manchón, Murphy, & Marin, 2008). As the transitions of different processes may signal writers’ strategy use (Levy & Ransdell, 1995), this study shows L2 writing benefits from efficient global planning at the beginning and focused production
during the translating process. Pedagogically, instructors need to cultivate L2 writers’ awareness of process management (see Mikulski & Elola, 2011) and strategy use (De Silva, 2015) to better scaffold their L2 writing development.

Conclusions and suggestions for future research

By analyzing how L2 writers of different writing skills paused in an argumentative writing task, this study has disclosed interesting pausing patterns that are insightful for comprehending the L2 writing process. First, writing skill does not affect time allocation or pausing patterns from a global perspective, testifying to the importance of qualitative use of task time. Second, writing skill influences interval pausing patterns, showing that writing products benefit from long planning pauses at the beginning of the writing. Finally, writing skill affects how writers manage their writing processes in terms of interaction and shifts between different component processes, with good writers displaying functionally well-defined writing stages while poor writers being forced to shift back and forth to sustain the writing. As effective writing demands competent process management and good strategy use (De Silva, 2015; Van den Bergh & Rijlaarsdam, 2001), L2 writers should strategically define clear goals for each writing stage to avoid draining their working memory capacity with multiple writing processes concurrently.

While good writing is an orchestration of good vocabulary, grammar, genre, topic and procedural knowledge, it is beneficial for L2 writers to consciously reflect on their writing processes and improve their management of such processes. Of course, it would oversimplify the issue by claiming that prolonged pauses contribute to text quality, but L2 writers need to be reassured that long pauses devoted to global planning lead to more fluent and productive writing processes. This positive understanding of pauses may help boost L2 writers’ confidence in striving for sufficient global planning and lift their anxiety in long episodes of pausing during writing.

There are a number of limitations in the study that merit caution for interpreting the findings or designing future studies. First, this study investigated the temporal distribution of pauses, leaving the issue of how pauses were qualitatively filled unaddressed. Second, as short and frequent pauses were observed during the first temporal interval for the less-skilled group, this suggests the division of the writing event into five intervals was not sufficient to depict the transition from the planning process to the translating process for this group. Further research into pauses in more writing intervals would be necessary to fine grain the observations made in this study.

Acknowledgements

The research reported in this paper was funded by the Youth Project of Humanities and Social Sciences, Ministry of Education, China (No. 14YJC740095) and the National Social Science Foundation of China (No. 16BYY099). We would like to thank the reviewers and ETS editors for their insightful suggestions that help improve the paper.

References


**Appendix**

<table>
<thead>
<tr>
<th>Points</th>
<th>Organization</th>
<th>Content</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No organization</td>
<td>Little idea</td>
<td>Incorrect spellings or a few isolated words</td>
</tr>
<tr>
<td>2</td>
<td>No focus; disorganized</td>
<td>Ideas random, inappropriate or illogical</td>
<td>Incomplete or incorrect sentences; severe errors that affect meaning</td>
</tr>
<tr>
<td>5</td>
<td>Attempts to focus; minimal organization</td>
<td>Ideas mixed; few transitions</td>
<td>Monotonous sentence structures; numerous errors that interfere with meaning</td>
</tr>
<tr>
<td>8</td>
<td>Single focus; some lapses or flaws in organization</td>
<td>Ideas not well supported or elaborated</td>
<td>Little variety in syntax; some evident errors</td>
</tr>
<tr>
<td>11</td>
<td>Single focus; logical organization</td>
<td>Ideas appropriate and varied</td>
<td>Varied sentence structure; few errors</td>
</tr>
<tr>
<td>14</td>
<td>Single, distinct focus; logical progression of ideas</td>
<td>Details effective, vivid, explicit and pertinent</td>
<td>Very few, if any, errors</td>
</tr>
</tbody>
</table>

*Note.* This rubric assigns each text into one of the levels above, and minor deviations from the description for each level except the “0” point level may result in the loss or gain of one point. Thus, a perfect score is 15 and the lowest score is zero.
The Effect of Socially Shared Regulation Approach on Learning Performance in Computer-Supported Collaborative Learning

Lanqin Zheng1*, Xin Li1 and Ronghuai Huang1,2

1School of Educational Technology, Faculty of Education, Beijing Normal University, Beijing, China // 2Smart Learning Institute, Beijing Normal University, Beijing, China // bnuzhenglq@bnu.edu.cn // lxbnu@mail.bnu.edu.cn // huangrh@bnu.edu.cn

*Corresponding author

(Submitted October 23, 2016; Revised February 10, 2017; Accepted April 28, 2017)

ABSTRACT

Students’ abilities to socially shared regulation of their learning are crucial to productive and successful collaborative learning. However, how group members sustain and regulate collaborative processes is a neglected area in the field of collaborative learning. Furthermore, how group members engage in socially shared regulation still remains to be resolved in the collaborative learning context. In this study, a socially shared regulation-embedded collaborative learning tool was developed to enable learners to collectively regulate their learning. Our study evaluated the effect of the socially shared regulation approach on learning performance. In total, 66 undergraduates were randomly assigned to two groups: experimental (socially shared regulation approach) and control (non-socially shared regulation approach). The results indicated that the proposed approach significantly improved participants’ learning achievements, group performance, and socially shared regulation frequency. Moreover, participants were satisfied with the proposed approach in terms of perceived usefulness, perceived ease of use, and cognitive load. The practical implications and future studies are also discussed based on the findings.

Keywords

Socially shared regulation, Computer-supported collaborative learning, Learning performance

Introduction

Over the past two decades, research on computer-supported collaborative learning (CSCL) have demonstrated positive learning outcomes. Most studies in this field centered on the quality of knowledge co-construction (Gan & Zhu, 2007; Kimmerle, Moskaliuk, & Cress, 2011). However, previous studies revealed that co-construction of knowledge was not easy to achieve (Kuhn, 2015) and true knowledge creation was rare (Siqin, van Aalst, & Chu, 2015). The major reason was that group members need to regulate their own and the whole groups’ learning processes (Järvelä et al., 2016). As we know, collaborative learning emphasizes co-construction of shared goals, shared task representations, and shared strategies in a shared space, otherwise without these shared components, this approach may create a divergent or less effective and satisfying space for learners (Järvelä & Hadwin, 2013). In order to achieve common goals and shared understanding, a group need to coordinate group efforts in an effective and efficient manner (Kwon, Liu, & Johnson, 2014).

It has been recognized that regulatory processes played a very critical role in collaborative learning settings (Rogat & Linnenbrink-Garcia, 2011; Volet, Summers, & Thurman, 2009). In CSCL contexts, group members’ engagement in interactions need to be facilitated by the group’s regulatory processes in the shared task space (Roschelle & Teasley, 1995). Constructing shared goals and strategies needs to be coordinated by group members’ motivations, cognitions, metacognitions, emotions, and behaviors. In addition, learning occurs in increasingly interactive manner (Isohätälä, Järvenoja, & Järvelä, 2017). Therefore, it was necessary to investigate regulatory processes beyond the individual (Hadwin, Järvelä, & Miller, 2011a). Socially shared regulation (SSR) emerged when group members worked together to regulate their cognitions, metacognitions, motivations, emotions, and behaviors in collaborative learning settings (Järvelä & Hadwin, 2013). Socially shared regulation was also transactive since it required individuals to collectively regulate collaborative learning processes (Hadwin, Järvelä, & Miller, 2011a). However, most learners were lack of joint regulatory skills when they completed collaborative learning tasks (Malmberg, Järvelä, Järvenoja, & Panadero, 2015). Furthermore, it was very difficult to collectively regulate cognitions, metacognitions, motivations, emotions, and behaviors because individuals were self-regulating agents (Järvelä, Volet, & Järvenoja, 2010). Thus, a large gap exists between the expectations and realities regarding socially shared regulation of learning in CSCL contexts.

CSCL technologies have been found to contribute to productive social interactions and construction of knowledge (Koschmann, 1996). However, most CSCL tools only support communication or sharing of information, rather than regulating collaborative learning processes (Gress & Hadwin, 2010). In this study, a
socially shared regulation-embedded CSCL tool was proposed and developed. To evaluate the effectiveness of the developed system, we examined the effects on learning achievements, group performance, socially shared regulation frequency, technology acceptance, and cognitive load. Students were divided into experimental and control groups. Those in the experimental group learned by using a socially shared regulation-embedded CSCL tool, while those in the control group learned by using a CSCL tool without a socially shared regulation mechanism. The following research questions were addressed to examine the effects of this new approach:

- Is there any significant difference in learning achievements between the experimental and control groups?
- Is there any significant difference in group performance between the experimental and control groups?
- Is there any significant difference in socially shared regulation frequency between the experimental and control groups?
- Is there any significant difference in perceived usefulness and perceived ease of use between the experimental and control groups?
- Is there any significant difference in cognitive load between the experimental and control groups?

**Literature review**

**Regulation of collaboration**

Regulation of collaboration is characterized as goal-directed metacognitive activities where group members take strategic control of their behavior, cognition, metacognition, motivation, and emotions through interactions (Hadwin et al., 2011a; Miller & Hadwin, 2015). Therefore, regulation differs from knowledge construction because the former focuses on constructing metacognition, meta-motivation, and meta-emotion, while the latter centers on constructing domain knowledge (Järvelä & Hadwin, 2013). Three types of regulation for successful collaboration exist: self-regulation, co-regulation, and socially shared regulation (Hadwin et al., 2011b; Järvelä & Hadwin, 2013).

Self-regulation of collaborative learning requires individuals to plan, monitor, evaluate, and adapt their motivations, cognitions, emotions, and behaviors during collaboration (Schunk & Zimmerman, 2008). Winne and Hadwin (1998; 2008) proposed four iterative and linked phases of self-regulated learning: defining the task, setting goals and planning how to reach them, enacting strategies, and adapting metacognition. Therefore, in the first phase, learners need to define the task as well as understand the task. Then the second phase sees them setting task goals and planning how to approach the task. In the third phase, learners apply tactics to achieve task goals as well as monitor and control the learning process. Finally, the fourth phase sees learners making adaptations to task perceptions, goals, plans, and strategies by self-evaluating their learning performance. During collaborative learning, self-regulating oneself is necessary and essential for successful collaboration. However, self-regulation alone is not enough because collaborating with others requires group members to be aware of one another’s motivations, cognitions, metacognitions, emotions, and behaviors, as well as to regulate each other’s goals, plans, and strategies.

Co-regulation of collaborative learning is also important for successful collaboration. In co-regulation, individuals socially regulate each other’s learning through questioning, prompting, and restating (Volet, Summers, & Thurman, 2009), and occurs when one group member guides, supports, or shapes others’ activities (Hadwin, Oshige, Gress, & Winne, 2010). Developing awareness of others’ goals and progress as well as monitoring and regulating others’ self-regulation are required (Miller & Hadwin, 2015). Therefore, each group member must keep track of one another’s progress in order to coordinate collaborative activities. To summarize, the aim of co-regulation is regulating one another’s self-regulation so as to provide services for the whole group.

**Socially shared regulation in collaborative learning**

Socially shared regulation refers to the process by which all group members regulate their motivations, cognitions, metacognitions, emotions, and behaviors to construct a shared outcome (Hadwin et al., 2011a). When all group members set goals, make plans, or monitor progress, they are engaged in socially shared regulation. Studies have found that students need to regulate motivations, cognitions, metacognitions, and emotions together during collaborative learning (Hurme, Merenluoto, & Järvelä, 2009; Järvelä & Järvenoja, 2011). Therefore, socially shared regulation centers on jointly coordinated activities working toward the same goal in collaborative learning.
Collaborative learning aims to co-construct shared task representations, shared goals, shared plans, and shared strategies. Therefore, these shared activities need to be leveraged through socially shared regulation in order to regulate everyone’s motivation, cognition, metacognition, emotion, and behavior. Socially shared regulation is a complex process that plays a vital role in collaborative learning (Kempler Rogat & Linnenbrink-Garcia, 2011). Group members must use socially shared regulation to orientate their tasks, set task goals, make plans, enact tactics, monitor their learning process, and evaluate their performance in order to achieve productive collaborative learning.

Despite the evidence, studies have demonstrated that learners failed to achieve socially shared regulation during collaborative learning (Kirschner & Erkens, 2013; Zimmerman & Schunk, 2011). This occurred because it was more difficult and complex to regulate at the group level than at the individual level (Winne, Hadwin, & Perry, 2013). In addition, learners were not equipped to regulate the collective activities and share group work (Kempler Rogat & Linnenbrink-Garcia, 2011). Thus, external support was essential for facilitating socially shared regulation during collaborative learning. Furthermore, Järvelä et al. (2015) posited that socially shared regulation can be leveraged and facilitated by technology-based tools. Winne (2015) also recommended software systems as a medium for regulating group work. However, some studies also revealed that use of technology-based tools in promoting socially shared regulation remained scarce (Järvelä & Hadwin, 2013; Järvelä et al., 2015). In addition, many efforts have been made to support cognitive activities in CSCL (Hmelo-Silver & Barrows, 2008; Saab, Joolingen, & Hout-Wolters, 2012; Zhang, Scardamalia, Reeve, & Messina, 2009). Therefore, research is limited regarding enhancing socially shared regulation in CSCL. The present study aims to bridge these identified gaps and develop the CSCL tool to support socially shared regulation in collaborative learning.

**Development of a socially shared regulation-embedded collaborative learning tool**

In order to promote socially shared regulation, a CSCL tool with a socially shared regulation approach was developed. This socially shared regulation-embedded CSCL tool facilitated socially shared regulation by group members jointly setting goals, making plans, selecting strategies, discussing online, monitoring learning processes, evaluating, and reflecting in a CSCL environment. In addition, some pop-up hints and prompts were incorporated to remind learners to collectively regulate during collaborative learning processes.

![Figure 1. Task evaluation](image)

This socially shared regulation-embedded CSCL tool includes seven modules. The first module addresses task perception, enabling learners to evaluate a task in terms of difficulty, value, and similarity with previous tasks (See Figure 1). Learners can input their prior knowledge so that other group members can be aware of peers’ prior knowledge. Learners can also download the task resources and check the task description. In the second module, each group member can set goals and make plans, whereas other group members can revise these goals and plans. However, learners usually need to jointly set goals, preset learning achievements, timelines, steps, and strategies after discussion (See Figure 2). Thus, when all group members agree upon the goals and plans, they can proceed onto the next steps. The third module supports learners in enacting strategies, including searching...
for information, making notes, and summarizing. The fourth module demonstrates the latest progress, including the amount of prior knowledge, number of postings, goals and plans, emotional status, current learning performance, and the number of enacting strategies (See Figure 3). Group members can monitor collaborative learning processes based on the latest progress to collectively regulate goals, plans, strategies, emotions, and behaviors. The fifth module supports group members in discussions online (See Figure 4). Here, when learners input negative emotion symbols, automatic pop-up system prompts encourage learners to be more positive. For example, if learners input a symbol indicating a sad emotion, the following pop-up appeared: “Don’t be too sad. Kindly discuss with your group members and you can find a solution. Be optimistic.” Eight kinds of emotions can be selected, including enjoyment, hope, pride, shame, anxiety, anger, hopelessness, and tiredness. The system prompts can facilitate regulation of emotions. The sixth module supports learners in submitting their group products online. As the submission date approaches, which was predetermined by each group, an automatic pop-up system prompt reminds learners of the deadline. Thus, learners can finish the collaborative learning task on time. The seventh module supports group members in reflecting and evaluating what they have learned as well as the whole collaborative learning process. If one group member believes they have not achieved a goal, our system will guide them to reset goals, revise plans, and study again. All of these modules were designed as technological interventions to facilitate socially shared regulation. Furthermore, Figure 5 demonstrates the associations between the technological interventions and SSR processes.

Figure 2. Setting goals and making plans

Figure 3. The statistics of strategies
Participants

In total, 66 undergraduate students who responded to study recruitment posters displayed at a university participated. They ranged in age from 19 to 21 years, with only three being male. They majored in psychology, educational science, or educational technology. All participants were randomly assigned into experimental and control groups. Each group included three undergraduates. Thus, 11 experimental groups conducted collaborative learning using a socially shared regulation-embedded CSCL tool, while 11 control groups learned using a CSCL tool without socially shared regulation. All participants had experienced collaborative learning in previous courses. However, three members of each group had never experienced collaborative learning together and only participated in this study once. Therefore, this was new experience for all participants.
Collaborative learning tasks

All groups completed a collaborative learning task closely related to the cultivation of abilities that was taught in their psychology course. The collaborative learning tasks of the experimental and control groups were identical. The group product was a Word document that indicated each group’s opinions and solutions to these problems. They collaboratively solved the following problems online:

- Alice is a good student with a remarkable academic record. Her teacher also believes that she is the best student in her class. But after graduation, Alice performs at only an average level in her career. In addition, her classmates who are not good at studies show better career performances than her. Please explain the phenomenon and analyze the reasons by using online discussion with your group members.

- “There are three-hundred and sixty trades, and every trade has its grand master”. This is an old Chinese saying. These grand masters can acquit themselves of various duties splendidly. Please analyze what kinds of abilities are crucial for these grand masters. Are these abilities inherent?

- Many people can win gold medals in the Olympic Games, but only a few people can win a Nobel Prize in China. Qian Xuesen, a famous scientist, once questioned, “Why innovative talents can’t be fostered in China?” Please analyze and explain the phenomena as well as propose your solutions.

Data collection and analysis

Data included a pre-test, a post-test, and questionnaires for measuring participants’ learning achievements, technology acceptance, and cognitive load. The present study assumed that socially shared regulation approach can improve learning achievements and group performance. The learning achievement was measured by the tests. The pre-test aimed to examine prior knowledge regarding collaborative learning tasks. It consisted of 10 single-choice items and five multiple-choice items, with a perfect score of 100. All of these items were associated with the concepts, characteristics, and the measurement of abilities. The post-test comprised five short answer questions and two open-ended questions, also with a perfect score of 100. These questions were related to the definitions, characteristics, individual differences of abilities as well as the methods to cultivation of abilities. The pre-test and post-test were developed by the experienced teachers. In addition, group performance was measured using the scores of group products. Moreover, the pre-test and post-test were independently assessed by the two raters. Cohen’s kappa was adopted to determine the inter-rater reliability coefficient. The kappa value for the pre-test and post-test were 0.91 and 0.86, respectively. Two raters also independently evaluated the group products. The Cohen’s kappa value achieved 0.89, indicating good reliability.

The technology acceptance questionnaire aimed to examine the perceived usefulness and perceived ease of use of our tool, which was adapted from the one developed by Chu, Hwang, Tsai, and Tseng (2010). It included seven items for “perceived ease of use”, and six items for “perceived usefulness.” The Cronbach’s alpha values were 0.75 and 0.92, respectively. The cognitive load questionnaire adapted from Lai and Hwang (2015) included eight items with a seven-point rating scheme. These items aimed to evaluate whether our tool increased the cognitive load of leaners. The Cronbach’s alpha value was 0.80.

A content analysis method was also adopted to analyze the frequency of socially shared regulation behaviors. The socially shared regulation episode was analyzed. The episode comprised pieces of dialogue with a shared focus and collective regulation of the activity (Grau & Whitebread, 2012). The socially shared regulation episode was identified and coded based on several criteria. First, the initiative dialogues that started the collective discussion were coded as one socially shared regulation episode. For example, group member A said: “Hello, everyone! Let’s start.” group member B said, “OK. Let’s begin discussing the first task.” Second, the dialogues that achieved the shared regulation of learning among all group members were also coded as one socially shared regulation episode. For example, group member A said: “Can we make a detailed plan first?” group member B said: “Yes, we should make a plan to specify how to achieve the goal.” group member C said: “OK, let’s make a plan now.” Third, only dialogues that reflected task perception, setting goals, making plans, enacting strategies, monitoring, reflecting, evaluating, and making adaptation were coded as socially shared regulation episodes. All discussion transcripts were independently coded by two raters so as to identify the socially shared regulation episodes. The Cohens’ kappa value achieved 0.81. All discrepancies were discussed and solved.
Experimental procedure

The experimental procedure is shown in Figure 6. Experimental and control groups completed a pre-test about prior knowledge. Then, during the collaborative learning activity, experimental group participants learned using a socially shared regulation-embedded CSCL tool and those in the control group learned using a CSCL tool without socially shared regulation. To conduct online collaboration using our tool, members of the same group were located in different laboratories. The experimental and control groups completed the same collaborative learning task for 2 hours. Next, all participants completed the post-test and the questionnaires to compare learning achievements, technology acceptance, and cognitive load. Finally, eight participants were randomly selected from the experimental group to conduct the interview so as to investigate their perceptions of learning using the new approach. The interviewees addressed the following questions:

- What are the advantages of this new approach (i.e., using a socially shared regulation-embedded CSCL tool to learn)? Why?
- Do you think this new approach (i.e., using a socially shared regulation-embedded CSCL tool to learn) is helpful for you in collaborative learning? Why?
- Would you recommend this new approach to your peers?

![Figure 6. The experimental procedure](image)

Results

Analysis of learning achievements

Before the experiment, all participants from the experimental and control groups took a pre-test to evaluate their prior knowledge. The respective mean values and standard deviations of the pre-test were 56.06 and 15.45 for the control group, and 51.97 and 13.40 for the experimental group. The t-test result demonstrated no significant differences between the two groups ($t = 1.149, p > .05$), indicating that the two groups had equivalent prior knowledge before the collaborative learning activity.

![Table 1. ANCOVA result of learning achievements on the post-test](table)

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>33</td>
<td>71.52</td>
<td>8.56</td>
<td>4.101*</td>
</tr>
<tr>
<td>Control group</td>
<td>33</td>
<td>66.33</td>
<td>10.03</td>
<td></td>
</tr>
</tbody>
</table>

Note. $p < .05$.

After the collaborative learning activity, experimental and control group post-test results were compared. Table 1 shows the analysis of covariance (ANCOVA) result of learning achievements. A positively significant difference was found between the experimental and control groups ($F = 4.101, p < .05$), implying that the socially shared
regulation approach helped improve students’ learning achievements. In addition, the mean of the experimental group was higher than that for the control group. Therefore, the socially shared regulation approach positively impacted on improving learning achievements.

**Analysis of group performance**

The group products of the experimental and control groups were analyzed to compare differences. As shown in Table 2, a significant difference in group performance existed between the experimental and control groups. This finding indicated that the socially shared regulation approach had a significant impact on group performance.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>33</td>
<td>89.27</td>
<td>2.78</td>
<td>3.35*</td>
</tr>
<tr>
<td>Control</td>
<td>33</td>
<td>86.58</td>
<td>3.69</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p < .01.

**Analysis of socially shared regulation frequency**

Online collaborative learning was a new experience for all group members. We were particularly interested in whether this new approach can promote socially shared regulation among group members. The frequencies of socially shared regulation between the experimental and control groups were analyzed and compared. As shown in Table 3, it was obvious that there were significant differences in socially shared regulation frequency between the experimental and control groups (t = 7.962, *p < .05*). Furthermore, compared with the control group, the experimental group had higher socially shared regulation frequency. This finding indicated that the socially shared regulation approach had a significant impact on promoting socially shared regulation in collaborative learning.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>33</td>
<td>20.73</td>
<td>4.60</td>
<td>7.962***</td>
</tr>
<tr>
<td>Control</td>
<td>33</td>
<td>7.90</td>
<td>2.70</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* ***p < .001.

**Analysis of perceived usefulness and perceived ease of use**

The technology acceptance included the perceived usefulness and perceived ease of use. As shown in Table 4, no difference was found regarding perceived usefulness between the experimental and control groups (t = 0.247, *p > .05*). No significant difference was also found with regard to perceived ease of use between the experimental and control groups (t = –0.125, *p > .05*). These findings implied that learners’ perceptions were not affected by the treatments.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usefulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>33</td>
<td>3.45</td>
<td>1.15</td>
<td>0.247</td>
</tr>
<tr>
<td>Control</td>
<td>33</td>
<td>3.38</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Perceived</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>33</td>
<td>4.41</td>
<td>0.77</td>
<td>-0.125</td>
</tr>
<tr>
<td>Control</td>
<td>33</td>
<td>4.43</td>
<td>0.63</td>
<td></td>
</tr>
</tbody>
</table>

**Analysis of cognitive load**

Because learning to use the socially shared regulation approach was a new experience for all participants, we also examined whether this new approach increased cognitive load during collaborative learning. Table 5 shows the t-test result of the cognitive load. No significant difference was found for cognitive load between the experimental and control groups (t = –0.773, *p > .05*). Therefore, the proposed socially shared regulation approach did not increase learners’ cognitive load.
Table 5. t-test result of cognitive load for experimental and control groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>33</td>
<td>3.02</td>
<td>0.81</td>
<td>-0.773</td>
</tr>
<tr>
<td>Control group</td>
<td>33</td>
<td>3.18</td>
<td>0.93</td>
<td></td>
</tr>
</tbody>
</table>

Interview results

To obtain a better understanding of learners’ perceptions of the collaborative learning activities, eight students from the experimental group were randomly selected for interviews. All interviewees reported one main difference between the collaborative learning activity and previous learning activity. That is, the socially shared regulation approach can promote awareness of collective regulation and facilitate knowledge building during collaborative learning. For example, student A indicated, “This tool can indeed improve the awareness of socially shared regulation. Consequently, our group performance was also improved.”

The results of interviews also revealed that learners believed that the socially shared regulation approach was helpful in facilitating setting of goals, making plans, monitoring the collaborative learning processes, and reflecting and evaluating learning outcomes and processes. For example, student B indicated, “This tool can facilitate setting goals and making plans. Usually, we are not aware of setting goals and making plans.” Student C stated, “This friendly tool reminds us to submit group products based on our schedule.” Student D said, “This tool can help us to reflect on the collaborative learning process and group products. This is very necessary for improving learning performance.”

All eight students reported that they would be willing to adopt our system using the socially shared regulation approach to conduct collaborative learning in the future. They also indicated that they would recommend the socially shared regulation approach to their peers.

Discussion and conclusions

We proposed a socially shared regulation approach for supporting collaborative learning activities by examining the learning achievements, group performance, socially shared regulation frequency, technology acceptance, and cognitive load between students situated in two different groups. One group worked in a socially shared regulation-embedded collaborative learning environment and the other in a conventional collaborative learning environment. We found that the socially shared regulation approach can significantly improve students’ learning achievements and group performance. This was consistent with Järvelä and Hadwin’s (2013) results, which demonstrated that socially shared regulation can promote collaborative knowledge building. This finding was also in line with Panadero and Järvelä (2015) who found that socially shared regulation improved learning outcomes. Our findings were also consistent with those of Mayordomo and Onrubia (2015), which revealed that there was an increase in collaborative knowledge construction when work group coordination was applied. Researchers have reported that the groups with the higher level of socially shared regulation achieved the higher group performance and learning achievements (Grau & Whitebread, 2012; Janssen, Erkens, Kirschner, & Kanselaar, 2012; Volet, Summers, & Thurman, 2009).

Furthermore, it was found that the proposed socially shared regulation approach can promote awareness and frequency of collective regulation. Figure 5 clearly demonstrated the relationships between the proposed interventions and socially shared regulatory processes. It was found that the functionalities provided by task perceptions informed users about other group members’ status, the availability of resources, and the difficulty of tasks so as to define tasks. The module of setting goals and making plans was effective for users to set learning goals and performance goals as well as make proper plans. Furthermore, Järvelä et al. (2015) also indicated that socially shared regulated learning environment should be designed to promote the activation of regulatory processes. Our tool provided the recommendation about learning strategies, which can help group members to select and enact strategies. In addition, online group discussion promoted the externalization of learning processes, which can facilitate the activation of joint regulation. Moreover, the usage information provided by the modules of history and the latest progress facilitated users to monitor the learning processes as well as increase the awareness of both self- and group members’ learning processes. The module of reflection and evaluation reminded learners to reflect whether they achieved the goals. If learners did not achieve the goals, our tool guide learners to iteratively regulator of the learning processes. Previous studies also revealed that regulated learning can be supported when the learning environment provided scaffolds (Van Merrienboer & Kirschner, 2013). Our socially shared regulation-embedded CSCL tool provided scaffold in the key regulatory
processes such as setting goals, making plans, enacting strategies, monitoring, and reflection to achieve a shared outcome. Consequently, the socially shared regulation can be achieved by increasing metacognitive awareness and regulatory at the group level.

In addition, this new approach did not negatively affect learners’ perceptions regarding usefulness and ease of use. Feedback from the experimental group revealed that they believed the socially shared regulation approach was useful and easy to use. In addition, the socially shared regulation approach did not increase learners’ cognitive load. Although there was no significant difference in cognitive load, the experimental group appeared lower than the control group. The major reason was that our system was friendly and not complex, thereby motivating students’ willingness to use it for an extended period. As Mayer and Moreno (2003) indicated that system designers should aim to reduce cognitive load and increase users’ pleasure by proper design as well as allocating cognitive resources.

In the field of CSCL, most studies focused on collaborative knowledge building among group members. Little attention was paid to socially shared regulation of task perceptions, goals, plans, and strategy use. In addition, Hadwin and Oshige (2011b) revealed that learners often experience problems in jointly regulating their learning. Therefore, the present study sought to improve group members’ socially shared regulatory skills during collaborative learning. The main contribution of this study lie in proposing and developing a collaborative learning tool using a socially shared regulation approach. The results of the empirical study validated the effectiveness of the proposed socially shared regulation approach.

The current study presents several implications for teachers and practitioners. First, socially shared regulation was a very important aspect for productive collaborative learning. Collaboration required all group members to successfully regulate their own learning and help other members to collectively regulate learning (Winne et al., 2013). Hence, teachers and practitioners should pay greater attention to socially shared regulation during collaborative learning. Teachers and practitioners were required to guide learners in collectively regulating beliefs, motivations, cognitions, metacognitions, emotions, and behaviors in order to achieve collaborative learning, rather than cooperative learning. For example, teachers encouraged group members to share responsibility before collaborative learning. Second, development or use of a technological tool can facilitate socially shared regulation during collaborative learning. Järvelä et al. (2015) reported on how these tools should increase group members’ awareness, promote activation of joint regulation of learning as well as support sharing and interactions with each other. Third, teachers should engage in collaborative learning to monitor learning processes and provide real-time feedback and intervention for learners.

This study has several limitations. First, the sample size was small, thereby making it difficult to generalize the results to other conditions. In future studies, we will enlarge the sample size and validate the approach in other contexts. Second, the collaborative learning tasks only focused on how to cultivate abilities in the field of psychology. Examining learners’ perceptions in other task contexts would be of value. Third, this study was conducted in a laboratory over a short period. Future studies should conduct a longitudinal study to examine the socially shared regulation approach in classrooms.

Acknowledgments

This study is supported by the youth project for pedagogy of the National Social Science Fund of China in 2014 “The empirical study on the assessment and intervention of adaptive regulation for college students” (CCA140154). This study is also funded by 2016 Comprehensive Discipline Construction Fund of Faculty of Education, Beijing Normal University.

References


Do Focused Self-Explanation Prompts Overcome Seductive Details? A Multimedia Study

Zhe Wang* and Olusola Adesope

Washington State University, Pullman, WA, USA // zhe.wang2@wsu.edu // olusola.adesope@wsu.edu

*Corresponding author

(Submitted April 22, 2016; Revised October 18, 2016; Accepted November 7, 2016)

ABSTRACT

Research on the seductive details effect on reading expository texts in multimedia learning environments has grown over the past few decades. However, less is known when seductive details are encountered in learning through worked-examples to solve problems. Thus, it is necessary to examine the seductive details effect when solving problems in a worked-example-based multimedia learning environment and the effect of focused self-explanation on seductive details. In the present experiment, the participants (N = 80) were randomly assigned to one of four different conditions and learned a multimedia lesson about electrical circuits on the computer: (a) seductive details, (b) seductive details and self-explanation prompts, (c) no seductive details, or (d) no seductive details but self-explanation prompts. Results showed that seductive details hindered learning because of increased extraneous cognitive load, and that focused self-explanation overcame the negative deleterious effect of seductive details through decreased extraneous cognitive load, and therefore improved learning. The theoretical and practical implications are discussed and future research directions are presented.

Keywords

Seductive details, Self-explanation, Multimedia learning, Worked examples

Introduction

Seductive details are materials that are interesting and entertaining but are irrelevant to the main ideas or learning objectives of a lesson (Harp & Mayer, 1997; Harp & Mayer, 1998). Rey (2012) systematically reviewed the seductive details effect literature and stated that overloading the working memory is one of the possible causes of the seductive details effect. Seductive details can be considered, by definition, as a source of extraneous cognitive load in that they are not relevant for learning and can be altered by instructional interventions (Beckmann, 2010; Park, Moreno, Seufert, & Brunken, 2011; Sweller, van Merrienboer, & Pass, 1998). Seductive details have been believed to trigger students’ situational interest through the natures of novelty and interestingness, which has been well documented recently (e.g., Magner, Schwanke, Alevan, Pospescu, & Renkl, 2014). However, by tempting students to spend their limited working memory resources in processing extraneous information, seductive details are damaging to the construction of a coherent mental representation (Park, Flowerday, & Brunken, 2015). Other than overloading the working memory, seductive details may hamper learning due to attention distraction (Lehman et al., 2007), coherence disruption (Mayer & Jackson, 2005), or schema interference (Harp & Mayer, 1998).

Solution to the seductive details effect

Technology has developed rapidly in education and offered a broad selection of multimedia tools for presenting complex learning information (e.g., Flikr, YoutTube). In spite of their easy accessibility and wide acceptability by learners, a main concern is that seductive details defined as interesting but irrelevant messages are hidden in these visually rich presentations. In the light of few information-picking choices learners could have when facing technology-based learning materials, to find a way to remedy the negative impact brought about by seductive details is of high importance. Unfortunately, regardless of abundant research asserting that seductive details are deleterious because they engender extraneous overload, results are yet inconclusive and hence no definite pragmatic solution has been found. Although a straightforward solution to the extraneous overload problem, suggested by researchers, is to eliminate words and pictures that are irrelevant to the instructional goal or at least insert these messages sparingly (Mayer & Fiorella, 2014; Mayer, Heiser, & Lonn, 2001; Rey, 2012), seductive details are expected everywhere in today’s learning environment particularly in computer-based presentations and thus simply excluding seductive details may not be feasible in many situations. Since seductive details are concerned with extraneous cognitive load, effective pedagogical approaches need to be developed to alleviate the load. Therefore, further research explicitly examining the role of different cognitive strategies (e.g., note-taking, making predictions, making inferences) in overcoming the seductive details effect can help the learner to deal with the complexity of multimedia learning environments.
A solution: Self-explanation

Up to date, the majority of research on seductive details has utilized expository texts as the instructional material. For instance, the topics used by previous studies include the process of lightning (e.g., Harp & Mayer, 1998; Lehman et al., 2007; Peshkam, Mensink, Putnam, & Rapp, 2011), what causes ice ages (Sanchez & Wiley, 2006), how a cold virus infects the human body (Mayer, Griffith, Jurkowitz, & Rothman, 2008), and the synthesis of ATP (Park et al., 2011; Park, Flowerday, & Brünken, 2015). In contrast, there are surprisingly few studies that focused on problem-solving domains (e.g., Towler et al., 2008). Problem-solving is a crucial component of learning competencies and is a complex process that requires domain-specific knowledge, structural knowledge, and metacognition (Bulu & Pedersen, 2012). Hence, this paucity presents an opportunity for research explicitly examining the seductive details effect in the process of solving problems. In the present study, the learning environment was designed through worked examples because they are regarded as supporting the initial acquisition of cognitive skills (Renkl, 2014) by applying the principle to problem solving. Furthermore, Renkl (2005) stated that providing worked examples is an effective method for problem-solving.

Although there is a common recognition of the importance of worked examples in skill acquisition, learning from worked examples does not guarantee favorable learning outcomes (Schworm & Renkl, 2006). Learners’ self-explanation activities are critical in fully exploiting the potential of instruction based on worked examples (Renkl, 2011). It is argued that the extent to which learners benefit from worked examples depends on how well they explain the provided solutions to themselves (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Renkl, Stark, Gruber, & Mandl, 1998).

Self-explanation refers to a reflective activity explaining to oneself the meanings of learning materials (Chi et al., 1989). Also, self-explanation can be considered as a product besides an activity. Ainsworth and Burcham (2007) pointed out that a self-explanation is additional knowledge generated by learners beyond the information they are given to study. Consider learners reading explanatory texts on ice ages for an example. The texts tell learners “CO₂ prevents long-wave radiation from escaping from the Earth into space.” The learner may ask “Why CO₂ is able to prevent such radiation” and then self-explain “It could be that CO₂ in the atmosphere is like a big sponge absorbing radiation.”

According to Chi (2000), self-explanation helps learners actively construct understanding in two ways: generating appropriate inferences and revising knowledge, which have been acknowledged by much research (e.g., Ainsworth & Burcham, 2007; Kwon, Kumalasari, & Howland, 2011), and recognized as two fundamental reflection mechanisms that can help learners engage in deeper sense-making activities, thus developing a deeper understanding of material (Rau, Aleven, & Rummel, 2015). In most cases, text is not expected to be a complete covering every detail, self-explanation can compensate for the inadequacy or ambiguity of the text by means of generating inferences. In addition, self-explanation can enable learners to realize a gap and compare their flawed mental models to those presented in the text, which can finally help them to revise their current models accordingly and resolve the dissonance.

Other than the benefits aforementioned, self-explanation may enhance learning by reducing extraneous cognitive load. Although engaging in the process of prompting open self-explanations may force learners to spend a great deal of extra time and effort in attending to task-extraneous aspects (e.g., where to start self-explaining; how deep a self-explanation should go), which usually require more working memory resources, certain forms of instructional assistance added to self-explanation are able to prevent working memory being overloaded. For example, providing a partial explanation could reduce extraneous cognitive load by reducing the size of the problem space (Kwon, Kumalasari, & Howland, 2011; Van Merriënboer & Sweller, 2005). Menu-based self-explanation prompts force essential and generative processing, while minimizing extraneous processing and maintaining motivation (Johnson & Mayer, 2010).

Therefore, in the manner of engaging in constructing inferences and assessing the gap, self-explanation may be able to moderate the diverting, disrupting, or distracting effect of seductive details. As McEldoon, Durkin, and Rittle-Johnson (2013) stated, self-explanation can benefit learning by focusing attention on relevant, underlying principles. In other words, learners are better able to distinguish between main concepts and secondary ideas during the self-explaining process and thus seductive details are processed at a minimum level or even sifted out. On the other hand, in learning through worked examples, prompts to self-explain specific solutions can reduce the element interactivity not essential to the task (Sweller, 2010), which may in turn compensate for the extraneous cognitive load caused by seductive details.
The present study

The experiment aimed to investigate the roles of focused self-explanation in learning from worked-out examples when seductive details are either included or excluded. Another focus of the study concerns the interactional relationship between self-explanation and seductive details on both extraneous cognitive load and transfer performance. The experiment was implemented in a computer-based multimedia learning environment and the learning material presented on computers was chosen from the field of physics focusing specifically on electrical engineering.

The study addressed four research questions: (a) are there any differences among the four different conditions in triggered situational interest, (b) are there any differences among the four groups in extraneous cognitive load, (c) are there any differences among the four groups in learning performance, and (d) are there any mediation effects of different conditions on learning performance through extraneous cognitive load?

Methods

Participants and design

Our sample consisted of 80 students in grade 10 (47 females, 33 males, average age = 16.6 years ranging from 15 to 18 years, $SD = .82$) from a middle school in China. Most of the participants had not studied electrical circuits systematically and intensively but had learned some basic facts (e.g., the difference between a series circuit and a parallel circuit). They were randomly assigned to one cell of a $2 \times 2$ between-subjects factorial design. The first factor was the presence or absence of seductive details. The second was the presence or absence of self-explanation prompts. Thus, there were 20 students in the seductive-details plus prompting condition (SP), 20 in the seductive-details only condition (SN), 20 in the no-seductive-details plus prompting condition (NP), and 20 in the no-seductive-details only condition: control condition (C).

Materials

In a computer-based learning environment, the multimedia instruction used in this study pertained to series and parallel circuits, particularly applying Ohm’s law in such circuits (e.g., determine the voltage if the current and resistance are known). The entire learning material was presented on 3 screens, the first one with a brief description of electrical circuits and Ohm’s law and its equations, the second one with a worked example related to a series circuit, and the third one with a worked example related to a parallel circuit. Specifically, the second screen contained a problem formulation, an equation, and solution steps (see Figure 1).

In the seductive-details conditions, interesting information presented in both image and text was inserted after the problem formulation (see Figure 2). The information was selected from online resources dealing with the topic of circuits but not relevant to the instructional goal. In the prompting conditions, each solution step was
followed by a self-explanation prompt (see Figure 3) with the purpose of encouraging the learners to self-explain the logic behind each solution step. The self-explanation prompts presented to the participants were focused, according to Wylie and Chi (2014), who argued that “focused self-explanation prompts provide more explicit instruction regarding what the content of the self-explanation should include” (p. 422).

Figure 2. Screenshot of the learning environment used in the seductive-details conditions

Figure 3. Screenshot of the learning environment used in the self-explanation prompting conditions

Measures

The pre-experimental questionnaire solicited general demographic information and also asked the participants to rate their knowledge of electrical circuits. The prior knowledge measure was a pre-test administered prior to the experimentation intended to test if participants’ existing knowledge was similar across different conditions. Specifically, prior knowledge was measured with five self-report questions (excellent internal consistency with Cronbach’s alpha = .93) about circuits (e.g., “How much do you know about electrical circuits?”) developed by one of the authors. The self-report questions asked the participants to rate their knowledge on a scale from 1 to 5 (1 being nothing and 5 being very much), yielding a possible score ranging from 5 to 25 points. Self-reported measures were considered a valid measure of prior knowledge and employed in several previous studies (e.g., Harp & Maslich, 2005; Harp & Mayer, 1998) regardless of the fact that it is often regarded as a subjective measure.
Triggered situational interest was assessed on a scale (good internal consistency with Cronbach’s alpha = .87) with 5 five-point Likert items (1 = strongly disagree, 5 = strongly agree) focusing on students’ spontaneous focused attention and affective reactions to the materials presented (e.g., “I think the material is interesting”). These items were adapted from a study by Flowerday and Schraw (2003). The triggered situational interest measure was included in order to examine the degree to which the seductive details used in the present study contributed to the arousal of interest.

Extraneous cognitive load was measured by three items that were adapted from the scale used by Park, Flowerday, and Brunken (2015). Each of the three items was rated on a five-point Likert scale (acceptable internal consistency with Cronbach’s alpha = .72) ranging from 1 to 7 (1 being extremely easy and 7 being extremely difficult). The items are associated with the unnecessary cognitive demands imposed by instructional design (e.g., “How easy or difficult did you find it to collect all information you needed in learning the worked examples?”).

Learning performance was assessed with a test containing 10 problems, which were not identical to the problems showed in the worked examples during the learning phase. Five problems assessed procedural knowledge in applying Ohm’s law in series circuits and the other five assessed procedural knowledge in applying Ohm’s law in parallel circuits. An example item is, “Resistors R1, R2, and R3 are plugged into a series circuit. The total voltage drop across R1, R2, and R3 is 24V. The resistance of resistors R2 and R3 is 2 Ω and 3 Ω respectively. The current through R1 is 4A. Question: What is the voltage drop across R1?” In each task, one point was assigned for each acceptable solution step written by each participant. Because the final solution to each problem could be achieved within five steps, the possible score for each participant ranged from 0 to 5 points. Five points were also given to an answer if the final solution had been reached with a correct answer provided even though one or two steps was omitted. Two independent raters trained with a mutually agreed scoring method scored each answer and reached an interrater reliability rating of $r = .92$.

Procedure

Upon arrival, participants were instructed to sit at one of the computers in a computer lab. Prior to launching the multimedia learning program, all the participants were asked to complete the pre-experimental questionnaire that includes the demographic survey and prior knowledge test. After they completed the questionnaire, the participants were randomly assigned to either SP, S, P, or C condition. All the participants were told to learn Ohm’s law in electrical circuits on computer. First of all, the description of Ohm’s law and features of electrical circuits were presented to the participant. Once the participant finished going through the prerequisite knowledge facts, he or she then proceeded to the next screen by clicking on a right arrow and learned through the example with solution steps provided. The participant then proceeded to the third screen where the second example was provided by clicking on a right arrow. The whole learning process is learner-paced. In other words, participants controlled the pace of the presentation by clicking a left or right arrow at the bottom of each screen to continue to the next screen or go back to the previous screen. The learning process was identical across the four conditions (SP, SN, NP, and C) except that those in the seductive-conditions received extra seductive image and text after the problem formulation was presented and those in the prompting conditions had to self-explain each solution step.

After participants learned the worked examples, the multimedia learning program was turned off and was no longer accessible to them. The participants were required to complete the triggered situational interest survey and cognitive load survey and then were presented with the 10 problems on the screen. They were asked to type solution steps and final answers to the problems on computer. Although working through the problems was self-paced, participants generally completed the questions within 40 minutes.

Results

Preliminary analyses

A one-way analysis of variance (ANOVA) was conducted to examine differences in prior knowledge among the conditions (SP, SN, NP, and C) before the intervention was implemented. The result indicated that the difference in the prior knowledge score among the conditions was not statistically significant ($F(3, 76) = .469, \text{MSE} = 3.512, p = .705; \eta^2 = .018$). Therefore, the conditions were considered fairly equivalent. Hence, a decision was
made not to use the prior knowledge score as a covariate in subsequent analyses. In addition, Table 1 shows the descriptive statistics for the triggered situational interest, extraneous cognitive load, and learning performance measures.

Table 1. Condition differences on extraneous cognitive load, triggered situational interest, and learning performance

<table>
<thead>
<tr>
<th>Condition</th>
<th>SP (n = 20)</th>
<th>SN (n = 20)</th>
<th>NP (n = 20)</th>
<th>C (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>TSI</td>
<td>18.2</td>
<td>3.14</td>
<td>17.7</td>
<td>2.99</td>
</tr>
<tr>
<td>ECL</td>
<td>7.75</td>
<td>2.34</td>
<td>9.5</td>
<td>1.73</td>
</tr>
<tr>
<td>LP</td>
<td>30.55</td>
<td>9.73</td>
<td>14.75</td>
<td>5.50</td>
</tr>
</tbody>
</table>

Note. SP = Seductive-details-prompting; SN = Seductive-details-no-prompting; NP = No-seductive-details-no-prompting; C = No-seductive-details-no-prompting; TSI = Triggered situational interest; ECL = Extraneous cognitive load; LP = Learning performance.

Triggered situational interest

An ANOVA was conducted to examine possible difference between the four different conditions on triggered situational interest and the result indicated that there was a significant difference ($F(3, 76) = 10.14, p < .001$) the seductive-details-promoting and seductive-details-no-prompting conditions scored similar ($p = .89$), while significantly higher than the no-seductive-details-promoting or no-seductive-details-no-prompting condition ($p < .05$), suggesting that the participants who read the seductive details were much more interested in the learning material as compared to those who did not.

Extraneous cognitive load

A 2 × 2 between-subjects analysis of variance (ANOVA) was conducted with prompting self-explanation (self-explanation prompts vs. no self-explanation prompts) and seductive details (seductive details vs. no seductive details) as the independent variables and extraneous cognitive load as the dependent variable. It showed a significant main effect of self-explanation ($F(1, 76) = 12.98, p < .001; \eta^2 = .146$) and a significant main effect of seductive details ($F(1, 76) = 16.26, p < .001; \eta^2 = .176$) but no significant interaction ($F(1, 76) = .026, p = .872, \eta^2 = .000$).

Post hoc Bonferroni tests detected significant differences for the seductive-details-no-prompting condition vs. no-seductive-details-no-prompting condition ($d = .91, p < .05$), seductive-details-prompting vs. no-seductive-details-prompting ($d = .89, p < .05$). In addition, there was a marginal difference between seductive-details-prompting and seductive-details-no-prompting ($d = .85, p = .057$). However, there was no significant difference between no-seductive-details-prompting and no-seductive-details-no-prompting ($d = .76, p > .05$).

Learning performance

A 2 × 2 between-subjects analysis of variance (ANOVA) was conducted with prompting self-explanation (self-explanation prompts vs. no self-explanation prompts) and seductive details (seductive details vs. no seductive details) as the independent variables and learning performance as the dependent variable. It showed a significant main effect of self-explanation ($F(1, 76) = 43.56, p < .001; \eta^2 = .364$) and a significant interaction effect between self-explanation and seductive details ($F(1, 76) = 7.14, p < .05; \eta^2 = .086$). Simple effects analysis demonstrated that there was a significant difference between seductive details and no seductive details at self-explanation ($p < .05$) but no significant difference at self-explanation ($p = .303$). Another simple effects analysis demonstrated that there was a significant difference between self-explanation and no self-explanation at seductive details ($p < .001$) and significant difference at no seductive details ($p < .05$).

Post hoc Bonferroni tests detected significant differences for the seductive-details-prompting condition vs. seductive-details-no-prompting condition ($d = 2.00, p < .001$), no-seductive-details-prompting vs. no-seductive-details-no-prompting ($d = .91, p < .05$), no-seductive-details-no-prompting vs. seductive-details-no-prompting ($d = 1.03, p < .05$). However, there was no significant difference between no-seductive-details-prompting and seductive-details-prompting ($d = -.29, p = .303$).
Mediation analyses on seductive-details condition

With a regression-based approach the mediation analysis helped to answer the questions about whether or not the seductive details effect/prompting effect is mediated by extraneous cognitive load. First, we were interested in a possible mediation effect of seductive details on learning performance through extraneous cognitive load in the absence of self-explanation prompts. Therefore, we tested whether seductive details (no-seductive-details-no-prompting vs. seductive-details-no-prompting) indirectly affected learning performance via extraneous cognitive load. Generally speaking, it is a valid model ($R^2 = .48, p < .001$). The analysis demonstrated that seductive details significantly increased extraneous cognitive load ($b = 1.95, SE = .68, p < .05$; path a in the mediation model). In addition, extraneous cognitive load significantly decreased performance, controlling for condition ($b = -1.74, SE = .40, p < .001$; path b in the mediation model). The analysis confirmed that there was an indirect effect of seductive details on learning performance ($a*b = -3.40, LCL = -7.18, UCL = -0.96$) via extraneous cognitive load. A visual representation of this mediation model can be found in Figure 4. Also, the results revealed the total effect ($b = -6.6, SE = 2.03, p < .05$) and the direct effect ($b = -3.19, SE = 1.84, p = .0912$) of seductive details on learning performance, which indicated the data were consistent with a complete mediation model.

![Mediation Model](image)

*Figure 4. Mediation results in the absence of self-explanation prompting*

We also tested whether extraneous cognitive load mediated the relation between seductive details and learning performance in the presence of self-explanation prompts. The no-seductive-details-prompting and seductive-details-prompting conditions were included in the analysis and the results revealed that seductive details significantly increased extraneous cognitive load ($b = 1.92, SE = .63, p < .05$). Extraneous cognitive load, however, did not significantly decrease learning performance, controlling for prior knowledge and condition ($b = -.88, SE = .71, p = .221$). Thus, there was no indirect effect of seductive details on learning performance ($a*b = -1.69, LCL = -5.68, UCL = .58$).

Mediation analyses on self-explanation condition

Given no significant difference between no-seductive-details-no-prompting and no-seductive-details-prompting on extraneous cognitive load, a mediation analysis was not necessary and it can be concluded that the relationship between prompting and learning performance was not mediated by extraneous cognitive load.

![Mediation Model](image)

*Figure 5. Mediation results in the presence of seductive details*

We also tested whether prompting indirectly affected learning performance via extraneous cognitive load in the presence of seductive details. The seductive-details-no-prompting and seductive-details-prompting conditions were included in the analysis and the model summary showed that it is a valid model ($R^2 = .74, p < .001$). The results revealed that prompting significantly decreased extraneous cognitive load ($b = -1.75, SE = .65, p < .05$). Extraneous cognitive load significantly decreased learning performance, controlling for prior knowledge and condition ($b = -2.64, SE = .46, p < .001$). Thus, there was an indirect effect of prompting on learning performance ($a*b = 4.62, LCL = 1.37, UCL = 8.93$) via extraneous cognitive load. A visual representation of this
mediation model can be found in Figure 5. Also, the results revealed the total effect ($b = 15.8$, $SE = 2.50$, $p < .001$) and the direct effect ($b = 11.18$, $SE = 2.00$, $p < .001$) of prompting on learning performance, which indicated the data were consistent with a partial mediation model.

Discussion

Differences in extraneous cognitive load

As expected, the seductive-details-no-prompting condition reported the highest extraneous cognitive load. This finding is consistent with previous studies (Mayer, Griffith, Jurkowitz, & Rothman, 2008; Sanchez & Wiley, 2006) confirming that seductive details constitute a significant source of extraneous cognitive load. The no-seductive-details-prompting condition reported the lowest extraneous cognitive load suggests that self-explanation prompts help learners concentrate on schema-relevant principles rather than direct attention to search processes / without searching in the vast information pool presented to them (Paas & van Gog, 2006; Sweller & Cooper, 1985). When the seductive-details condition was prompted to self-explain, they reported moderate levels of extraneous cognitive load, as well as the no-seductive-details-no-prompting condition, suggesting that self-explanation was able to substantially relieve the load caused by the seductive details while not reaching the most desirable level.

The role of self-explanation prompting

An important finding of the study is that when learners who received seductive details were prompted to self-explain the worked example, the seductive details effect was reduced substantially. As evidenced by the indirect effect of prompting on learning performance via decreased extraneous cognitive load from the mediation analysis, we safely conclude that self-explanation is an effective learning strategy that can direct learners' cognitive resources to the construction of problem-solving schema underlying each solution of the worked example, whereas at the same time cognitive processing capacity is not drawn by irrelevant messages, rendering seductive details processed minimally. As Fiorella and Mayer (2015) pointed out, self-explaining, as a generative learning activity, can prime the cognitive processes of selecting, organizing, and integrating, and thus leaves cognitive resources for schema acquisition rather than seductive details.

Another crucial result relates to the comparison between the no-seductive-details-no-prompting and no-seductive-details-prompting conditions. We found that, while seductive details were excluded, prompting enhanced learning but, surprisingly, not through reduced extraneous cognitive load, which is supported by the mediation analysis. Indeed, learning without seductive details does not introduce a source of extraneous load that is needed to be reduced by self-explanation prompts. Better learning outcome associated with the prompting condition is consistent with previous work that has proved that the generation of self-explanations is an effective cognitive activity that yields germane cognitive load and enhances learning (Chi et al., 1989; Renkl, 1997; Renkl & Atkinson, 2002). According to the active processing principle of Cognitive Theory of Multimedia Learning (CTML: Mayer, 2014), learners may not learn most effectively unless instruction methods aim at generative processing whose level is represented by germane cognitive load.

When extraneous cognitive load fails to explain

We also found, unexpectedly, that the learning performance was not in perfect correspondence with the extraneous cognitive load for another comparison. That is, regardless of the higher levels of extraneous cognitive load for the seductive-details-prompting condition versus the no-seductive-details-prompting condition, they performed similarly and outperformed the rest two conditions. One reasonable explanation is that extraneous cognitive load is not the only psychological construct that affects learning. Indeed, increased triggered situational interest for the seductive-details-prompting condition may compensate for some of the negative effect of extraneous cognitive load. An alternative explanation is that it is not always the case that the lower the extraneous cognitive load, the better the learning outcomes. The classic cognitive load theory (CLT: Paas, Renkl, & Sweller, 2003; Plass, Moreno, & Brunken, 2010; Sweller, van Merrienboer, & Paas, 1998; van Merrienboer & Sweller, 2005) infers that free cognitive processing capacity is the difference between the working memory resources and the total load consisting of intrinsic, extraneous, and germane load. Therefore, as long as the load does not exceed the working memory capacity, the level of the extraneous cognitive load is of little concern (intrinsic load is not considered here because it is irreducible by instructional design and germane load is...
believed to remain constant because both conditions employed prompting). In the present study, the extraneous load of 7.75 for the seductive-details-prompting condition and that of 5.95 for the no-seductive-details-prompting condition might not reach a critical number, above which learning would be hurt.

Conclusions and limitations

The results of this study add significantly to the existing body of evidence showing that including seductive details in an expository text is detrimental to learning performance. Specifically, inserting seductive details to the lesson format of worked examples leads to reduced learning outcomes when compared with materials without any seductive details. As a result, an instructor might simply forgo seductive details when designing worked examples to minimize the student’s risk of limited working memory being overloaded. Furthermore, the study attests to the benefit of applying self-explanation prompts in learning when seductive details are received as part of learning materials. Hence, instructors can be encouraged to prompt their students to self-explain crucial elements and/or interactivity of elements covered by a lesson as a way of primarily focusing their cognitive resources on the construction of schema rather than encoding irrelevant information. We also note that it is recommended to design prompts that are focused versus open for beginning learners, because focused self-explanation prompts enable them to proceed under guidance, avoiding another source of load that could have been introduced by open prompts.

On the basis of this study investigating a specific way of presenting the material (seductive details) and a particular instructional technique (self-explanation prompts), our findings can extend and provide empirical evidence for a potential broader interpretation of the relation between cognitive load and learning. We contend that the relationship between extraneous processing and learning outcomes is not linear all the way. As observed in the present study, as long as the total load did not reach a warning level, there would be of less need for reducing extraneous load. One educational implication that can be drawn is that if an instructor is concerned about non-essential material causing a modest amount of extraneous processing, he or she is encouraged to not only reduce extraneous load but also pay attention to seeking learning activities that can promote generative processing to increase germane load.

The results of this study have important educational implications considering the known effect of self-explanation on seductive details. Teachers might incorporate oral self-explanation prompts when illustrating important conceptual and/or procedural skills. Textbook writers and multimedia designers should insert visual self-explanation prompts appropriately in a lesson. Furthermore, future research should focus on seeking and investigating other instructional techniques to reduce extraneous cognitive load caused by processing seductive details. With findings provided by studies on other generative or constructive learning activities, we can make a strong case for how to effectively overcome the seductive details effect in multimedia learning.

There are some limitations in conducting this study. One of the major limitations is that we did not record learners’ oral self-explanations or collect their written self-explanations. Without analyzing their oral or written explanations, little is known about the quantity and/or quality of self-explanations they made, which would limit the potential of self-explanation prompts. Second, using self-report scales measuring extraneous cognitive load could be problematic with validity because participants tend to be inaccurate in estimating the load they have experienced relying upon introspection. Future research is needed to examine the objective measures of cognitive load, such as response time to a secondary task during learning proposed by DeLeeuw and Mayer (2008) and the rhythm method suggested by Park and Brunken (2015).

The present research made contributions by means of moving from merely examining the seductive details effect in current literature to seeking pedagogical techniques to overcome seductive details. Given the moderation effect of self-explaining we have observed in this study, we encourage researchers to explore other techniques particularly generative learning activities such as imaging, drawing, and mapping to overcome the seductive details effect.

References


Enhancing Students’ Computer Programming Performances, Critical Thinking Awareness and Attitudes towards Programming: An Online Peer-Assessment Attempt

Xiao-Ming Wang¹, Gwo-Jen Hwang²*, Zi-Yun Liang² and Hsiu-Ying Wang³

¹School of Teacher Education, Zhejiang Normal University, Jinhua, China // ²Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan, R.O.C. // ³Department of Education, National University of Tainan, Tainan, Taiwan, R.O.C. // zgjhwang.academic@gmail.com // ghwang.academic@gmail.com // flora91410@gmail.com // cute@tkgsh.tn.edu.tw

*Corresponding author

(Submitted May 21, 2016; Revised September 21, 2016; Accepted December 11, 2016)

ABSTRACT

It has become an important and challenging issue to foster students’ concepts and skills of computer programming. Scholars believe that programming training could promote students’ higher order thinking performance; however, many school teachers have reported the difficulty of teaching programming courses. Although several previous studies have attempted to develop friendly user interfaces to ease students’ loads, teaching programming courses remains a big challenge for most school teachers. In this study, an online peer assessment-based system was developed to cope with this problem. The students could use the peer-assessment function to provide comments to peers, and review the feedback and scores from peers during the learning activity. A quasi experiment was conducted on four classes of 166 ninth graders of a junior high school located in southern Taiwan to examine the impacts of the developed system. Two classes of students were assigned to the experimental group, learning with an online peer assessment-based teaching strategy, while the other two classes were the control group, learning with the conventional teaching strategy. The experimental results showed that the students in the experimental group had better programming knowledge and skills as well as more positive learning attitudes and critical thinking awareness than those in the control group, revealing the benefits of the proposed approach.

Keywords

Computer programming, Peer assessment, Critical thinking, Learning attitude, Scratch

Introduction

How to effectively conduct programming education to help students develop the concepts and skills of programming has become an important and challenging issue (Kran, Mladenović, & Rosić, 2015; Yang et al., 2015; Brito & de Sá-Soares, 2014). Many researchers believe that engaging in programming tasks not only enables students to gain knowledge of programming, but also has great potential in promoting their higher order thinking performance (Keppens & Hay, 2008; Wang, Huang, & Hwang, 2016; Williams et al., 2002). However, past experience shows that many students perceive learning computer programming as a difficult and boring task (Kátai, 2015). Therefore, researchers have attempted to develop programming environments with graphical interfaces to ease students’ load of developing computer programs. Scratch is such a programming environment that has been widely adopted by school teachers. According to the report of Kobsiripat (2015), who used Scratch as the programming language to let students produce their own works, it was found that the students’ performance was increased due to the easy manipulation learning environment in Scratch.

Despite the advancements in programming tools and environments, teaching programming languages remains a big challenge for most school teachers (Barr & Guzdial, 2015; Sáez-López, Román-González, & Vázquez-Cano, 2016). Researchers have suggested that, to effectively promote students’ learning performance in programming courses, it is necessary and important to include proper teaching strategies (Chang, Wu, Weng, & Sung, 2012). As computer programming is relevant to not only programming knowledge (i.e., knowing how a computer program works and the syntax of a programming language), but also programming skills (i.e., knowing how to analyze a task and develop a program for dealing with the task), it could be more effective to engage students in the practice of viewing and emulating with guidance from the teacher. Project-based learning with peer assessment is such an approach that engages students in playing the roles of project developer and reviewer based on the rubrics provided by the teacher. Such a learning activity provides great opportunities to create bigger learning effects for students than the conventional teaching and practice approach (Gielen, Peeters, Dochy, Ongena, & Struyven, 2010; Hsia, Huang, & Hwang, 2016; Liu & Carless, 2006). The value that peer assessment brings to education is that in the process of peer assessment, students have opportunities to learn the rubrics provided by the teacher in depth, and to learn from peers’ work since they play the role of reviewer. This
implies that they are situated in the context of practicing critical thinking about peers’ work and providing concrete suggestions and feedback to their peers (Hovardas, Tsivitanidou, & Zacharia, 2014). In addition, while evaluating others’ work, they make comparisons of their own work with that of others; hence, they are likely to ponder whether their work meets the teacher’s standards. Such a learning process provides a strong chance of developing students’ critical thinking awareness (Boud, Cohen, & Sampson, 1999; Liu & Carless, 2006; Topping, 1998; Xiao & Lucking, 2008).

On the other hand, several studies have also reported that peer assessment might put pressure on students due to the interpersonal relationships involved, which might lead to negative impacts (Falchikov, 2004). Vanderhoven, Raes, Schellens, and Montrieux (2012) suggested that online peer assessment could be a solution to cope with this problem. Compared to traditional face-to-face peer assessment, using computer technologies that enable students to provide feedback to peers anonymously could significantly reduce their pressure and avoid the negative impacts. However, little research has attempted to employ the online peer assessment approach for programming projects. Moreover, to the best of our knowledge, no empirical study has been conducted to show the effects of online peer assessment on students’ programming knowledge or skills. Therefore, in this study, an online peer assessment system was developed and an experiment was conducted to examine students’ computer programming performance using the system and the Scratch programming environment. The research questions of this study are as follows:

- Can an online peer assessment-based approach increase students’ learning achievement (i.e., programming knowledge and skills) in Scratch programming?
- Can an online peer assessment-based approach increase students’ learning attitude toward Scratch programming?
- Can an online peer assessment-based approach increase students’ awareness of critical thinking?

Literature review

Scratch programming

Scratch is a programming language that was developed by the Lifelong Kindergarten research group of the Massachusetts Institute of Technology in 2007. Scratch was designed to help young people learn creative thinking, logistical inferencing, and collaborative learning, which are considered to be basic skills needed in the 21st century. Scratch provides users with a visual interface which allows them to join blocks together to easily create interactive stories, games, and animations on their own. It uses blocks as the objects to command. Such an innovative programming environment means beginners can avoid making syntax and logical errors (Ouahbi, Kaddari, Darhmaoui, Elachqar, & Lahmine, 2015). This method allows students to create multimedia works easily and develop their creativity and thinking ability during the learning process (Resnick et al., 2009; Wang, Huang, & Hwang, 2016). Many scholars believe that Scratch is an innovative learning method for students to be creative in this digital era (Kim & Song, 2012; Moreno, 2012; Peppler & Kafai, 2007).

In recent years, an increasing number of tools using Scratch as a programming language have been adopted in junior high and elementary schools. In Ouahbi, Kaddari, Darhmaoui, Elachqar, and Lahmine’s (2015) experiment in one senior high school, it was found that using Scratch to let students program can bring out their potential for developing games; 65% of them expressed that they were willing to continue using Scratch to create, while only 10.3% of the students were willing to use the traditional Pascal programming language to program. The conclusion reached was that using Scratch to let students learn programming languages in a basic programming design course can effectively increase their learning motivation; students do not need to worry about the programming syntax, and can easily develop their games and animations. Meanwhile, many researchers have pointed out that using Scratch to teach programming can let students learn the syntax for judging and looping, and for creating their own personal works (Fesakis & Serafeim, 2009; Peppler & Kafai, 2005). Kobsiripat (2015) adopted Scratch to conduct a programming design learning activity for 60 elementary school students; it was found that through this activity, the students’ creativity and higher order thinking benefited. Wang, Huang, and Hwang (2016) conducted a project-based learning activity for students with advanced and ordinary performance using the Scratch programming language; the results showed that the advanced students performed better in terms of their learning achievement, problem-solving ability, learning attitude, and learning motivation than the ordinary students.

In this study, Scratch was adopted as the programming tool owing to the following reasons. First, it provides a graphical programming interface which enables novice learners and young students to realize the basic programming logic without being confused by programming language syntax. Moreover, it has been widely
adopted for a number of years now in Taiwan’s high schools, including the selected school, as the learning tool in programming courses.

**Peer assessment**

Peer assessment, as proposed by Topping (1998), is carefully designed guidance for giving feedback to peers, and can be seen as a teaching strategy to help students review the advantages and weaknesses of their peers’ work so that they can make amendments to the goals not reached while developing their metacognitive ability, critical thinking awareness, test performance, and professional skills (Joordens, Pare, & Pruesse, 2009; Topping, 2009). Peer assessment can be applied as a formative or summative evaluation by providing students with rubrics from teachers to evaluate their peers’ work and to give suggestions (Liu & Carless, 2006; Topping, 1998). During the peer assessment process, students need to view works from the teacher’s point of view and consider whether their peers’ performance meets the teacher’s requirements. On the other hand, while evaluating peers’ work, students are also given a chance to reflect on and think about how to improve their work; such a process can benefit both reviewers’ and reviewees’ learning performance, and promote students’ creativity (Boud, Cohen, & Sampson, 1999; Xiao & Lucking, 2008).

Topping and Ehly (2001) believed that peer assessment is one of the most important ways of peer-assisted learning, providing students with a way to learn from each other, which is beneficial for developing knowledge and skills. While students are evaluating their peers’ work, they can imitate the strengths of the work, a view which happens to coincide with that of Schunk (1987), while having students with similar abilities and background learn together and imitate each other brings positive learning effects. Tsivitanidou, Zacharia, and Hovardas (2011) discovered that peer assessment learning activities can encourage positive learning attitudes on the part of students. Tseng and Tsai (2010) further pointed out that using online peer assessment has the advantage of being anonymous, and students do not need to worry that they will be identified while giving scores, which encourages them to raise doubts and propose the advantages and drawbacks of others’ work to achieve a better learning effect. Davies (2000) used a computer system to administer a peer assessment activity, and it was found that after comparing their work with that of others, the students had higher self-awareness and demands, which has a positive meaning for education. Hwang, Hung, and Chen (2014) adopted a quasi-experiment in natural science courses in which the elementary school students were asked to design games; the results showed that with the peer assessment mechanism, the students’ learning achievement, learning motivation, and problem-solving ability were effectively increased. Khonbi and Sadeghi (2012) found that students learning with peer assessment performed better than those with self-assessment.

To date, several studies have confirmed that peer assessment in most learning activities is an effective and reliable learning approach (Falchikov & Magin, 1997; Khonbi & Sadeghi, 2012). Students can ponder the advantages and weaknesses of peers’ work before submitting their feedback; such a process provides students with a powerful chance to reflect and to assist their peers in producing work of better quality (Tsai & Chuang, 2013; Cheng, Liang, & Tsai, 2015). It also provides the opportunity to improve their critical thinking, which refers to the cognitive process, a logistical conclusion or solution gained from the judgement of learning goals (Glassner, Weinstock, & Neuman, 2005; van Gelder, 2005).

As students in peer assessment activities have a chance to reflect on their own work and that of their peers using the rubrics provided by the teacher, their critical thinking awareness would be increased, especially when using online peer assessment tools (Joordens, Pare, & Pruesse, 2009; Lynch, McNamara, & Seery, 2012). In peer assessment activities, students can review their learning process and their results by evaluating others’ work and accepting others’ feedback, helping them to find their own learning blind spots and reconstruct their learning goals and plans (Cheong & Cheung, 2008). On the other hand, many researchers have further pointed out that using computer technology to learn can cultivate students’ critical thinking awareness (Dwyer, Hogan, & Stewart, 2014; Wang, Huang, & Hwang, 2016). Therefore, in this study, an online peer-assessment strategy is employed in a Scratch programming activity.

**Online peer-assessment system for computer programming**

To evaluate the effectiveness of the proposed teaching strategy, an online peer assessment system was developed and provided for the experimental group students to conduct peer assessment, as shown in Figure 1. There are three databases in the peer-assessment environment: (1) a “work” database for students to save their Scratch programming works, student personal information, and peer assessment database; (2) a “personal information”
database to keep the students’ profiles, such as their names, student IDs and contact information; and (3) a “peer-assessment” database to save the ratings and comments provided by the students and teachers.

In addition, there are several function units that assist students and teachers during the peer-assessment process: (1) a “Do project” function to enable students to upload their programming projects; (2) a peer-assessment system to support assessment activities by providing an anonymous matching mechanism (i.e., the function unit that automatically assigns students’ work to their peers), an instant reminder mechanism (i.e., the function unit to inform the students to review peers’ work), and an instant feedback mechanism (i.e., the function unit to inform the students to browse the ratings and comments from peers); and (3) a “Monitor system” for teachers to supervise the peer-assessment process.

![Databases](image1)

**Figure 1.** Online peer assessment system for computer programming projects

Figure 2 shows the user interface of the peer-assessment system. When reviewing peers’ work on a computer screen, the students can rate the work and provide comments on the other screen.

![User interface](image2)

**Figure 2.** Screenshot of peer assessment in the online peer assessment system
When the students submit their scores and feedback, a reminder window is displayed for the reviewers to check their comments. Besides, another anonymous comment is given once the reviewer submits his/her comments, as shown in Figure 3. Furthermore, the system also offers the teacher the function to monitor the status of peer assessment.

![Ratings and comments from peers](image1)

**Figure 3. Review peers’ comments on their work**

**Experimental design**

**Participants**

The participants of this study were from four classes of 166 ninth graders in a junior high school located in southern Taiwan who learned Scratch programming in their computer courses. A quasi-experimental design was adopted, and two classes of 80 students were assigned to the experimental group, learning with the online anonymous peer assessment teaching strategy. Every student gave scores and comments for the work assigned by the system. The other two classes of 86 students were assigned to the control group, learning with the traditional teacher feedback teaching strategy. The teaching activity was led by an experienced teacher with over 30 years of teaching experience in information technology. After the activity, the students were asked to make a Scratch programming on the topic of school life or on a social issue of their choice.

**Experimental procedure**

The experiment took 10 weeks of two hours per week, as shown in Figure 4. In the first week, the students took a prior knowledge test about the Scratch programming language and completed the pre-questionnaires of learning attitude and critical thinking. In weeks 2-5, the teacher taught how to program for four weeks, and the students were asked to develop a program as the programming skills pre-test. Following that, one class period was spent on the introduction of the programming design project and the rubrics of the work.

In weeks 6 and 7, the students did the programming project. In weeks 8 and 9, the students in the experimental group were situated in the online peer assessment activity, including evaluating peers’ work based on the rubrics provided by the teacher (see Table 1) and revising their work after receiving ratings and comments from peers. On the other hand, those in the control group received comments from the teachers and revised their work accordingly. It should be noted that the lowest rating in the rubrics was “1” rather than “0” to avoid overly discouraging the reviewees. Similar rating schemes have been adopted by several previous studies (Hsia et al., 2016; Hwang, Hung, & Chen, 2014).

In order to examine the effectiveness of the proposed teaching strategy, in the tenth week, all of the students were asked to finish a Scratch program as the post-test of programming skills; following that, they took the
programming knowledge post-test and completed the questionnaires of learning attitude and critical thinking awareness.

166 ninth graders

The experimental group (N=80)  The control group (N=86)

Programming knowledge pre-test
Pre-questionnaires of learning attitude and critical thinking

Instruction of Scratch programming
programming skill pre-test

Programming project
Online peer-assessment feedback
Revise the project

Programming project
Traditional teacher feedback
Revise the project

Programming knowledge and programming skill post-tests
post-questionnaires of learning attitude and critical thinking

week 1

weeks 2-5

weeks 8-9

week 10

Figure 4. Experimental process

Table 1. Rubrics of online peer assessment

<table>
<thead>
<tr>
<th>Dimension</th>
<th>4 points (Excellent)</th>
<th>3 points (Good)</th>
<th>2 points (Fair)</th>
<th>1 point (Poor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant level of topics and contents</td>
<td>The content and topic are highly relevant. There is bountiful content with detailed descriptions.</td>
<td>Most of the content is relevant to the topic but it can be more bountiful.</td>
<td>Little content in the project is relevant to the topic and there should be more content.</td>
<td>The content and the topic are totally irrelevant.</td>
</tr>
<tr>
<td>Level of integration of multimedia elements and content</td>
<td>High integration of pictures, videos, and contexts.</td>
<td>Most of the integration of pictures, videos, and contexts is great.</td>
<td>There is little integration of pictures, videos, and contexts.</td>
<td>No integration of pictures, videos, and contexts.</td>
</tr>
<tr>
<td>Joyfulness and innovation of contents</td>
<td>The project is presented vividly with multiple characters.</td>
<td>Most of the content is fun and there are many characters, but only a few have special features.</td>
<td>Only one character is presented in the project, but the character has special features.</td>
<td>The content is boring with only one character with no features.</td>
</tr>
<tr>
<td>Gaming scene design</td>
<td>There are more than three scenes in the project.</td>
<td>There are three scenes in the project.</td>
<td>There are two scenes in the project.</td>
<td>There is only one scene, making the game monotonous.</td>
</tr>
</tbody>
</table>

Measuring tools

To achieve the research goal, the measuring tools in this study included the programming knowledge tests, programming skill test, and the questionnaires of learning attitude and critical thinking.
The programming knowledge tests included a pre-test and a post-test designed by two teachers who had taught the computer technology course for over 15 years. The pre-test aimed to test the students’ Scratch prior knowledge, using 15 true-or-false test items and 20 multiple-choice test items, with a perfect score of 100. The post-test aimed to test the students’ knowledge of Scratch programming, such as the meaning and syntax of variables, array, conditional branch, loop statements, input, output and the control operations, using 10 true-or-false test items, 15 multiple-choice test items, and 5 fill-in-the-blank test items, with a perfect score of 100. The KR20 values of the programming knowledge pre-test and post-test were 0.71 and 0.73, respectively. As both the values were greater than 0.7, it is concluded that the reliability of the tests was good according to the suggestions of Nunnally (1978).

The programming skill pre-test and post-test were two programming projects, which aimed to evaluate the students’ competence of using the programming statements and operations to develop Scratch programs based on the topics specified by the teacher. The program was evaluated by two teachers based on the rubrics in Table 1. The consistency degrees (i.e., Pearson correlation coefficients) of the ratings given by the two teachers were 0.95 and 0.92 for the programming skill pre-test and post-test, respectively.

The learning attitude questionnaire was developed by Hwang, Yang, and Wang (2013). There were nine items and the Cronbach’s alpha value was .79 in this study, showing that the reliability of the questionnaire was acceptable.

The critical thinking awareness questionnaire was revised based on the measure developed by Dwyer, Hogan and Stewart (2014). It aimed to measure the students’ tendency to review and reflect on their learning status and the problems they were facing, such as “I periodically examine if I have achieved the learning objectives” and “When facing a problem, I would question if there are better choices or solutions.” The Cronbach’s alpha value of the questionnaire was .85 in this study, showing that the questionnaire was of good reliability.

Results

Analysis of programming knowledge and skills

The programming knowledge pre-test aimed to test if the students had equivalent programming background knowledge, while the post-test aimed to evaluate their programming knowledge after the learning activity. On the other hand, the Scratch programming pre-test was used to assess the students’ programming skills before the learning activity, while the Scratch programming post-test was used to evaluate their programming skills after the activity. The Scratch programs developed by the students were assessed by two experienced teachers to ensure the objectives of the evaluation.

A test for the homogeneity of regression on the programming knowledge pre-test scores was conducted and it was found that there was no significant difference between the scores of programming knowledge ($F = 0.818, p = .367$) and programming skills ($F = 3.038, p = .083$) of the two groups, meaning that they had equivalent programming knowledge and skills before the activity.

<p>| Table 2. ANCOVA result of the learning achievement post-test and programming skills of the two groups |
|-------------------------------|-------------------------------|-------------------------------|-------------------------------|----------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Variance</th>
<th>Group</th>
<th>$N$</th>
<th>Mean</th>
<th>$SD$</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge test</td>
<td>Experimental</td>
<td>80</td>
<td>74.1</td>
<td>10.5</td>
<td>73.8</td>
<td>1.37</td>
<td>29.92***</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>86</td>
<td>62.9</td>
<td>14.8</td>
<td>63.3</td>
<td>1.32</td>
<td></td>
</tr>
<tr>
<td>Scratch project</td>
<td>Experimental</td>
<td>80</td>
<td>83.57</td>
<td>10.38</td>
<td>83.23</td>
<td>0.47</td>
<td>35.83***</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>86</td>
<td>73.38</td>
<td>10.94</td>
<td>73.70</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>

Note. *** $p < .001$.

An analysis of covariance (ANCOVA) was then conducted on the students’ post-test scores of programming knowledge and skills by adopting their pre-test scores of programming knowledge and skills as the covariate variables, respectively. As shown in Table 2, the experimental group, with online peer assessment teaching strategy outperformed the control group, with traditional teacher feedback. For the programming knowledge test, there was a significant difference between the scores of the two groups ($F = 29.92, p < .001$), and the experimental group had a higher score mean. Moreover, for the Scratch program, the experimental group also performed better than the control group ($F = 35.83, p < .001$). Such results indicate that the online peer-assessment approach brought better learning effects on students’ programming knowledge and skills than the conventional teaching.
Analysis of critical thinking awareness

To explore the effects of the proposed teaching strategy on students’ critical thinking awareness, ANCOVA was used to analyze the students’ post-questionnaire ratings by excluding the impacts of their pre-questionnaire ratings. An ANOVA test for the homogeneity of regression on the pre-questionnaire of critical thinking awareness was conducted, and no significant difference was found between the two groups of students for critical thinking awareness ($F = 1.90, p > .05$). An analysis of covariance was performed on the post-questionnaire scores, and the results showed that with different teaching strategies, a significant difference was reached for critical thinking awareness ($F = 21.23, p < .001$), as shown in Table 3. These results indicate that an online peer assessment approach can reinforce students’ critical thinking awareness.

<table>
<thead>
<tr>
<th>Variance</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical thinking</td>
<td>Experimental</td>
<td>80</td>
<td>3.78</td>
<td>0.67</td>
<td>3.80</td>
<td>0.07</td>
<td>21.23***</td>
</tr>
<tr>
<td>awareness</td>
<td>Control</td>
<td>86</td>
<td>3.38</td>
<td>0.67</td>
<td>3.36</td>
<td>0.07</td>
<td></td>
</tr>
</tbody>
</table>

Note. ***$p < .001$.

Analysis of learning attitudes

To explore the effects of the proposed teaching strategy on the students’ learning attitudes, ANCOVA was used to analyze the students’ post-questionnaire ratings by excluding the impacts of their pre-questionnaire ratings. A test for the homogeneity of regression on the pre-questionnaire of learning attitude was conducted, and no significant difference was found between the two groups of students for learning attitude ($F = 3.20, p > 0.05$), meaning that the two groups of students had similar levels of learning attitude. An analysis of covariance was performed on the post-questionnaire scores, and the results showed that with different teaching strategies, a significant difference was reached for the students’ learning attitudes ($F = 19.53, p < 0.001$), as shown in Table 4.

<table>
<thead>
<tr>
<th>Variance</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning attitude</td>
<td>Experimental</td>
<td>80</td>
<td>3.84</td>
<td>0.67</td>
<td>3.88</td>
<td>0.06</td>
<td>19.53***</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>86</td>
<td>3.52</td>
<td>0.73</td>
<td>3.49</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

Note. ***$p < .001$.

Discussion and conclusions

In this study, an online peer-assessment approach for computer programming activities was proposed, and a peer-assessment system was developed accordingly. A learning activity was conducted in a high school Scratch programming course to evaluate the effectiveness of the proposed approach. After analyzing the experimental data, the researchers of this study found that the students participating in the programming projects with the online peer assessment approach outperformed those participating in the projects with the conventional teaching and practicing approach in terms of programming knowledge, programming skills, critical thinking awareness, and learning attitude.

As indicated by a number of researchers, students face several challenges when learning computer programming knowledge and skills (Sáez-López, Román-González, & Vázquez-Cano, 2016). One challenge is the difficulty for novices to realize the meaning and syntax of a programming language which is very different from those of the natural language they use. Another challenge is to foster students’ competence of developing computer programs based on the programming knowledge they have learned. The adoption of visualized programming environments such as Scratch might solve part of these problems since students do not need to write programming statements using a language they are not familiar with. Instead, they can focus on understanding the meaning of the programming instructions and on learning to express their programming logic. In the meantime, the lead-in of peer-assessment could further guide them to realize the criteria of developing a computer program and to reflect on their programming project based on their observations of peers’ work as well as the feedback from their peers (Khonbi & Sadeghi, 2012; Tseng & Tsai, 2010; Vanderhoven, Raes, Schellens, & Montrieus, 2012).

Several previous peer-assessment studies have reported that deep involvement in learning projects by playing the role of a reviewer is able to encourage students to learn better (Hsia, Huang, & Hwang, 2016; Xiao & Lucking,
When playing the role of a reviewer, students are guided to think from the teacher’s point of view by referring to the rubrics provided by the teacher. Such a process is helpful to them for realizing the criteria of a successful work, as well as encouraging them to elaborate their work, which could promote their knowledge and skills of completing the learning tasks (Boud, Cohen, & Sampson, 1999; Lai & Hwang, 2015). Researchers have also indicated that the feature of anonymity in online peer assessment can encourage students to provide more comments on peers’ work with less hesitation (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010).

As indicated by Wang, Hwang, and Hwang (2016), the peer-assessment activity is a process of critical thinking. Students need to state the advantages and weaknesses of peers’ work and their own after playing the role of a reviewer. This is beneficial to them in terms of promoting their critical thinking awareness. In the online peer-assessment environment, the students were able to ponder on and revise their comments before sending any messages to their peers; that is, the online peer-assessment environment provided opportunities for the students to elaborate their comments and evaluate their peers’ work by thinking critically and deeply, confirming what was reported in the study of Joordens, Pare, and Pruesse (2009).

Consequently, in the Scratch programming environment with the peer-assessment approach, students are likely to have deep involvement, see things from different perspectives, and be more willing to provide comments, which could also be the reasons why they showed better programming knowledge and skills, learning attitude toward the programming course, and awareness of critical thinking than those who learned in the Scratch programming environment without the peer-assessment approach.

There is, however, one limitation of this study; that is, the findings might not be able to be inferred to those computer programming activities that use traditional programming environments with text interfaces instead of visualized environments like Scratch. This implies that further studies are required to investigate the effects of online peer assessment on students’ programming performance in different programming environments.

Acknowledgements

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

References


Students’ Reactions to Different Levels of Game Scenarios: A Cognitive Style Approach

Zhi-Hong Chen1∗, Sherry Y. Chen2 and Chih-Hao Chien3

1Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan //
2Graduate Institute of Learning Technology, National Central University, Taiwan // 3Department of Information Communication, Yuan Ze University, Taiwan // zhchen@ntnu.edu.tw // sherry@cl.ncu.edu.tw // b.jian1020@gmail.com

∗Corresponding author

(Submitted March 24, 2016; Revised June 10, 2016; Accepted January 13, 2017)

ABSTRACT

Game-based learning comprises of a set of concrete scenarios, which can be presented with various presentation modalities (e.g., text, text with graphic, and context). Such presentation modalities disseminate information in different ways. On the other hand, cognitive styles affect how people process information. Accordingly, the study presented in this paper investigates the influences of cognitive styles on students’ performance within different levels of game-based scenarios. More specifically, this study examines how students with different cognitive styles (i.e., Holist/Serialist) react to three presentation modalities (i.e., text, text with graphic, and context) in game-based scenarios. The results of an empirical study with 96 students revealed that all students made improvement, regardless of any presentation modalities. However, Holists significantly performed better than Serialists in the context version. According to the results, further suggestions for system design are discussed.

Keywords

Cognitive styles, Game-based learning, Situated contexts

Introduction

With the advance of multimedia technology, the use of multimedia in education has been regarded as a promising approach to optimizing student learning. Such advocacy is underpinned by the theory of multimedia learning (Mayer, 2002), which involves an assumption of how people learn from words, sounds, and images (Mayer & Moreno, 2003). More specifically, multimedia learning asserts that optimal learning occurs when visual and verbal information is presented together simultaneously. This is due to the fact that students have separate channels to process visual and verbal information. Accordingly, students have more opportunities to receive integrative information when visual and verbal forms are linked together. By doing so, students’ comprehension (Rusanganwa, 2013) can be enhanced and information can also be stored in the long-term memory structure (Kulhavy, Stock, & Kealy, 1993).

In spite of such benefits, it is still difficult for students to apply what they have learned to real situations because multimedia learning highlights the integration of multiple representation channels, rather than authentic transferable examples. To facilitate such learning transfer, scenario-based learning (Clarke, & Mayer, 2011; Clark, 2009; Kindley, 2002) is proposed based on the principles of situated learning theory (Lave & Wenger, 1991), which argues that learning should be situated in a specific context, or embedded in a particular social and physical environment (Kindley, 2002). Scenario-based learning makes students immune in authentic works, and then integrates needed knowledge and skills in the context, instead of abstract or decontextualized knowledge (Clarke, & Mayer, 2011). In other words, scenario-based learning advocates that students should learn in concrete situations and by examples.

In short, scenario-based learning highlights the significance of real situations, which provides three major potential benefits: engaging, meaningful, and transfer learning. Regarding engaging learning, unlike abstract information, information presented in an authentic way can stimulate students to be engaged in the learning process (Herrington, Reeves & Oliver, 2014), where different scaffolding designs can also be applied to enhance their engagement. Regarding meaningful learning, such a scenario could offer students rich information, including objects themselves as well as the relationships among various objects. Such rich information might help students understand how to organize them in an appropriate way so students can undertake meaningful learning. Regarding transfer learning, scenarios could trigger students to propose questions and acquire knowledge with concrete examples. When students learn from a number of different examples, they could synthesize and analyze similarities and differences among such examples. This experience can help them conduct and deduce what they have learned. Thus, they can know better how to apply their knowledge and skills.
from one scenario to similar scenarios in the future (Cormier & Hagman, 2014; McKeough, Lupart, & Marini, 2013).

Due to such benefits, the scenario has been widely applied in different learning settings, such as game-based learning and technology-enhanced language learning. On the one hand, game-based learning is often represented as a virtual world, which comprises of a set of concrete scenarios, to allow students to explore every place or interact with each other (Chien et al., 2013; Toscano et al., 2015). For instance, Barab and his colleagues (2005) developed a game scenario, which was designed as a virtual island to facilitate students’ scientific inquiry. In the virtual island, students were guided by a set of quests to observe, propose and evaluate their hypotheses. Such exploration and interaction in digital games are beneficial to student learning, especially in motivation and learning performance (Kebricht, Hirumi, & Bai, 2010; Papastergiou, 2009; Tsai, Yu, & Hsiao, 2012; Tüzün et al., 2009). On the other hand, technology-enhanced language learning often involves different scenarios, where technologies are used as tools to engage or enrich students’ learning experience. For instance, Di Blas and Paolini (2014) created a 3D multi-user learning environment to promote students’ language learning. Students are situated in a scenario, where they can control their avatars to interact with other students or communicate with NPCs (Non-player-characters) to practice their foreign language (Ibáñez et al., 2011).

The aforementioned studies deliver information within scenarios, which could be categorized as three different levels: text, graphic, and context. The three levels of the scenarios vary in how much information conveyed to students and how such information is presented. The text reveals symbolic and abstract information whereas the graphic offers visual and concrete information. The context not only delivers rich information about individual object, but also their detailed relationships among these objects. Accordingly, students have to process various types of information simultaneously in the scenarios (Kalyuga & Plass, 2009). On the other hand, not all of students can effectively cope with such multiple information sources (Ishii & Yamauchi, 1994). Therefore, individual differences should be considered in scenario-based learning.

Among various individual differences, cognitive styles particularly represent consistent individual different preferences of organizing and processing information and experience (Messick, 1976). Cognitive styles refer to the way of how students think, perceive, and remember information, or their preferred approaches to using such information (Riding & Grimley, 1999). For instance, Pask’s Holism and Serialism are one of significant classifications of cognitive styles. More specifically, Holists prefer to process information in a “whole-to-part” sequence whereas Serialists favor a “part-to-whole” processing of information (Jonassen & Grabowski, 2012). Recently, differences between Holists and Serialists are investigated in several studies. Specifically, it has been found that Holists and Serialists had different preferences for their navigational styles (Clewley et al., 2011). The former tended to take a non-linear pattern by “jumping” between different levels of subject contents with hypertext links. Conversely, the latter preferred to follow a linear pattern by having a suggested route or looking at the subject content step-by-step with back/forward buttons. The differences between Holists and Serialists are also found in their reaction to collaborative learning (Chen & Chang, 2016), where there is a need to provide either Holists or Serialists with additional help because heterogeneous groups with both Holists and Serialists can obtain the best learning performance. In addition, to meet the different needs of Holists and Serialists, an adaptive hypermedia learning system was developed to enhance student learning (Mampadi et al., 2011). The result showed that such an adaptive hypermedia learning system had positive effects on students’ perceptions and performance.

Although the aforementioned studies demonstrate fruitful results, there is a lack of studies to investigate how Holists and Serialists react to different levels of scenarios in the context of a game-based learning system. This issue is significant because it can provide concrete guidance on how to accommodate the needs of different cognitive style groups. To this end, this study aims to fill this gap. More specifically, the research question of this study is: How do Holists and Serialists react to different levels of scenarios in game-based learning systems? To answer the question, an empirical study was conducted. The details are described in the following sections.

**Methodology design**

**Development of a game-based learning system**

As mentioned in the earlier section, to facilitate learning transfer, digital games with different scenarios are paid enough attention in the area of learning technologies. Accordingly, we developed a game-based learning system, which taught English vocabularies on the topics of life and leisure, including the scenarios of a living room, bedroom, bathroom and kitchen. The goal of the game is to nurture a cartoon monster to grow up by satisfying...
all of its needs in life, such as eating, sleeping, and playing. The monster would tell the student objects that are needed so he/she is required to go to a specific scenario (e.g., living room, bedroom, bathroom, or kitchen) to look for such objects. For instance, if the monster is hungry, the student is required to go to the kitchen to cook. Then, the student can learn related foods (e.g., beansprout, carrot, chicken breast) in such a scenario. In other words, a scenario-based learning approach was adopted in this game, where students learn English vocabularies in an authentic scenario and then integrates what they have learnt to satisfy the needs of the monster.

To do comprehensive investigation to answer the research question of this research, the scenario of the game-based learning system was divided into three different levels: text, text with graphic, and a whole context. Thus, three versions were produced to correspond to the aforementioned three levels. More specifically, the three versions have some features in common.

- **Audio elements:** Audio was considered in all of three versions, which can provide the pronunciation of each vocabulary and its Chinese explanation because audio is one of most common elements in multimedia presentations (Clarke & Mayer, 2016) as well as important to vocabulary learning.
- **Non-linear learning:** The game-based learning system presents vocabularies in a non-sequential format. In other words, the learning system does not limit the orders of movements or learning paths, and allows the students to freely choose vocabularies by themselves.

However, the three versions differ in representational levels (i.e., text, text with graphic, and context). The first version (the text version) presented vocabularies in the form of text only. This version offered a basic setting when compared with the other two versions. The second version (text with graphic version) offered students additional explanations in the form of both text and graphic. The third version (context version) presented comprehensive information, including text and graphic explanations of the vocabulary as well as the surrounding objects to form as a meaningful learning environment. In addition, the text version and the text with graphic version presented vocabularies in an alphabetical order, whereas the context version presented vocabularies based on the arrangement of objects displayed in a scenario. Table 1 summarizes the common and different features among the three versions.

<table>
<thead>
<tr>
<th></th>
<th>Text version</th>
<th>Text with graphic version</th>
<th>Context version</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common features:</strong></td>
<td>To show vocabularies with pronunciation</td>
<td>To present the Chinese explanations of vocabularies</td>
<td>To display vocabularies in a non-linear format</td>
</tr>
<tr>
<td><strong>Different features:</strong></td>
<td>To show vocabularies in the format of text only (one channel)</td>
<td>To show vocabulary in the formats of text and graphic (two channels)</td>
<td>To list vocabularies based on the arrangement of a scenario</td>
</tr>
</tbody>
</table>

**Instruments**

- **Study preferences questionnaire:** Among various instruments, the Study Preferences Questionnaire (SPQ) can do a quick and easy measure of Holists and Serialist biases. Additionally, the SPQ had been used in several studies (e.g., Clewley, Chen, & Liu, 2011) and showed adequate reliability ($\alpha = 0.67$) in past research (Mampadi, Chen, Ghinea, & Chen, 2011). Consequently, this study adopted the Chinese version of the SPQ to identify Holists and Serialists and followed the criteria suggested by the original producer (Ford, 1985) to identify Holists and Serialists. More specifically, the participants were provided with two sets of 17
statements and they were requested to choose the statements they agreed or to indicate no preferences. Based on their choices, if they agree with over half of the statements related to Holists, they are identified as Holists. Conversely, they are considered as Serialists.

- **Achievement test:** The pre-test and post-test were employed to evaluate the participants’ levels of knowledge of the subject domain (i.e., English vocabulary). Each test consisted of 20 items, which asked the participants to fill out appropriate vocabularies according to the given sentences and such vocabularies were introduced in the system presented in the section of “development of a game-based learning system”. To enhance the reliability, same vocabularies were applied in both of the tests but sentences that presented the vocabularies in the pre-test were rephrased in the post-test. More specifically, the test item used in the pre-test, e.g., “An _____ is a round vegetable with many white layers on its inside which have a strong, sharp smell and taste” was initially developed by a college teacher and then the other college teacher rephrased this sentence as “An _____ is a round vegetable with many white layers on its inside which have a strong, sharp smell and taste” in the post-test. By doing so, the test items used in these two tests were compatible in order that the reliability of the results could be enhanced. Regardless of the pre-test or post-test, each test item had 5 point, and the total scores of a test ranged from 0 to 100.

**Participants and procedure**

To answer the aforementioned research question, a between-subjects quasi-experimental design was conducted. In total, three classes participated in the experiment and they were randomly assigned to the three groups, each of which used one of three levels of game scenarios. Regardless of any groups, all participants (N = 96) were second-year students in a University in Taiwan and they had basic computing skills and Internet experience. More specifically, the first group is control group one (CG1), where 37 participants used the text version; the second group is control group two (CG2), where 27 participants used the text with graphic version; the third group is the experimental group (EG), where 32 participants used the context version. During the experiment, each group needed to follow the procedures below:

- Identification of cognitive styles: All participants of the three groups were requested to fill out the SPQ. According to the results of the SPQ, the participants were furthermore classified into Holists and Serialists. The details are illustrated in Table 2.
- Conduction of the pre-test: The participants were asked to take the pre-test to identify their prior knowledge.
- Brief introduction: Before using the game-based learning systems, the participants were given a 10-minute introduction to the functionality of the game-based learning system assigned to this group.
- Interaction with the game-based learning system: To make students pay attention to the features of subject domain in the game-based learning system, they were asked to do the task sheet. More specifically, the participants were required to learn English vocabularies and then to complete sentences by filling out appropriate vocabularies in the task sheet, which included 20 fill-in-the-blank questions. By doing so, the participants needed to put effort to find those answers from the game-based learning system. In addition, to avoid the topic bias, the questions were designed to cover various topics, including accommodation, food and clothing. Regardless of the system versions, 40 minutes were allocated for each participant to complete the tasks.
- Conduction of the post-test: The participants were further requested to take the post-test to identify their learning performance.

<table>
<thead>
<tr>
<th>Table 2. Settings of the three groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text version (CG1)</td>
</tr>
<tr>
<td>Basic computing skill</td>
</tr>
<tr>
<td>Internet experience</td>
</tr>
<tr>
<td>Holists/Serialists</td>
</tr>
</tbody>
</table>

**Data analysis**

The independent variables of the experiment were the game-based learning systems (with three levels of game scenarios) and cognitive styles (with two different styles) whereas the dependent variable was score obtained from the pre-/post-tests. Since the independent variables involves two factors (i.e., game scenarios and cognitive styles), a two-way ANOVA was conducted to examine whether interactions occur between these two factors. Then, the analysis of scores was further carried out from two aspects: within-group and all-groups.
• **Within-group aspect:** Paired-samples t-test for each group was conducted to examine the difference between the pre-test and post-test on each cognitive style group. In addition, one-way analysis of covariance (ANCOVA) for each group was further carried out to examine the difference between Holists and Serialists, with cognitive styles as a between-subject variable and pre-test scores as a covariate.

• **All-groups aspect:** One-way analysis of variances (ANOVA) for each cognitive style was conducted to examine the difference among the three groups, using groups as a between-subject variable. All these analyses were conducted with a Statistical Package for the Social Science (SPSS v20).

**Results and discussion**

To examine whether interactions occur between these two factors (i.e., game scenarios and cognitive styles), a two-way ANOVA was first conducted. The result showed that no interactions exist \((F_{1,2} = 2.315, p > .05)\) between the two factors. It implies that the two factors only had simple main effects on the students’ achievement test. Consequently, the analysis of scores was further carried out from two aspects, i.e., within-group and all-groups.

**Holists and Serialists within each group**

Table 3 displays the means and SDs for the pre-/post-test of each group. The results from the paired-sample t-tests indicated that the scores of the post-tests in all groups were all significantly higher than the pre-test scores in these groups, regardless of Holists or Serialists. It implied that students with all of the versions could make significant improvement. This might be because all of the three versions had text and audio, which presented visual and verbal information together simultaneously. As suggested by the dual-coding theory, visual and verbal cues are processed differently to create separate representations (Paivio, 1971). Information coded by the two channels can increase the retrieval efficiency when compared with the information coded in a single channel. In other words, combining two channels, i.e., visual channel and verbal channel, could work better than using one channel only (Clarke & Mayer, 2016).

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Text version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holists</td>
<td>29.81</td>
<td>18.42</td>
<td>49.26</td>
</tr>
<tr>
<td>Serialists</td>
<td>28.50</td>
<td>15.10</td>
<td>47.50</td>
</tr>
<tr>
<td>Text with graphic version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holists</td>
<td>28.67</td>
<td>18.94</td>
<td>47.33</td>
</tr>
<tr>
<td>Serialists</td>
<td>29.58</td>
<td>11.17</td>
<td>49.16</td>
</tr>
<tr>
<td>Context version</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holists</td>
<td>22.64</td>
<td>12.38</td>
<td>43.23</td>
</tr>
<tr>
<td>Serialists</td>
<td>34.00</td>
<td>20.19</td>
<td>43.33</td>
</tr>
</tbody>
</table>

*Note. **p < .01.*

Furthermore, the one-way ANCOVA was applied to examine whether a significant difference existed between Holists and Serialists within each group, in terms of their improved scores (Table 4). The results indicated that there was a significant difference \((F_{1} = 8.181, p < .01)\) within the context version but such a significant difference was not found within the other two versions. After examining the means difference between Holists and Serialists within the context version (20.59 and 9.33), we found that Holists gained two times scores more than Serialists. In other words, Holists could get more benefits from the context version than Serialists.

<table>
<thead>
<tr>
<th></th>
<th>Holists</th>
<th>Serialists</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Text version</td>
<td>19.44</td>
<td>12.73</td>
<td>19.00</td>
</tr>
<tr>
<td>Text with graphic version</td>
<td>18.67</td>
<td>14.69</td>
<td>19.58</td>
</tr>
<tr>
<td>Context version</td>
<td>20.59</td>
<td>9.98</td>
<td>9.33</td>
</tr>
</tbody>
</table>

*Note. **p < .01.*

In short, the results seemed to imply that all of the three versions could foster both Holists and Serialists to improve their learning performance and that the context version was more beneficial to Holists than Serialists. A possible explanation was that Holists preferred to process information in a “whole-to-part” sequence (Jonassen
& Grabowski, 2012). The context version not only provided text and graphic explanations of the vocabulary, but also displayed the surrounding situation, which offered Holists a global picture to guide their vocabulary learning. Thus, the context version matched with Holists’ preferences so that they could make great improvement.

On the other hand, the context version might not match with the preferences of Serialists, who preferred to follow a “part-to-whole” sequence. More specifically, the context version provided multiple vocabularies in a situation so that Serialists might feel distracted. In other words, the context version might make Serialists have difficulties in identifying each individual vocabulary. This might be due to the fact that they tended to use a predominantly local learning approach, concentrating on one thing at a time (Ford, 1995). Because of such a local learning approach, they did not leap great improvement in the context version.

### Holists and Serialists for all groups

Table 5 displays the means and SDs of the improved scores of Holists and Serialists for all groups. To examine how Holists react to the three versions, an ANOVA was further conducted. The results showed that no significance existed among the three groups \((F_{(2)} = 0.096, p > .05)\). It implied that Holists showed similar improved scores in the three versions. A possible reason was that all of the three versions displayed information in a non-sequential format so Holists were allowed to use a non-linear approach to restructure information. Such an approach matched with internally directed characteristics that Holists possess (Mampadi, Chen, Ghinea, & Chen, 2011). More specifically, Holists are good at re-organizing information to adjust themselves to suit to existing environments (Clewley, Chen, & Liu, 2011), regardless of the text version, text with graphic version or context version. This may be the reason why Holists performed similarly in these three versions.

<table>
<thead>
<tr>
<th></th>
<th>Text version</th>
<th>Text with graphic version</th>
<th>Context version</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holists</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>19.44</td>
<td>12.73</td>
<td>18.67</td>
<td>14.69</td>
</tr>
<tr>
<td>Serialists</td>
<td>19.00</td>
<td>8.43</td>
<td>19.58</td>
<td>8.38</td>
</tr>
</tbody>
</table>

Note. **p < .01.

Unlike Holists, the results from an ANOVA demonstrated that Serialists showed significantly different reactions to the three versions \((F_{(2)} = 6.568, p < .01)\), and a further Tukey HSD post-hoc test revealed that Serialists with the text version and the text with graphic version significantly obtained higher improved scores than those with the context version. It seemed to imply that the text version and the text with graphic version, instead of the context version, were useful for Serialists to improve their learning performance. In other words, the context version seemed to be inappropriate to Serialists.

A possible explanation is that the context version offered massive information in the same situation at a time. Such rich information was not appreciated by Serialists, who preferred to concentrate on one thing at a time (Ford, 1995). In short, the context version did not match with their preferences of “part-to-whole” sequence that Serialists possessed. Conversely, the text version, and the text with graphic version offered a predominantly local learning sequence—one vocabulary at a time, which supported the needs of Serialists. Therefore, Serialists could gain great improvement in the text version and the text with graphic version, rather than the context version.

### Implication for system design

According to the aforementioned results, three implications for system design could be discussed here. The first implication is about why CG1 (i.e., text version) and CG2 (i.e., text with graphic version) can accommodate the preferences of Holists and Serialists. In CG1 and CG2, all of vocabularies are presented so that Holists can obtain a global picture of the subject content. Consequently, such a way matches with the preferences of Holists so their scores can be significantly improved. Additionally, all of vocabularies presented in CG1 and CG2 are shown in an alphabetical order, which supports Serialists’ sequential approaches. In addition to showing these vocabularies (i.e., imagery channel), their pronunciation (i.e., verbal channel) are also offered. According to the dual-coding theory (Clark & Paivio, 1991), their achievements can be significantly improved. This might be the reason why both of CG1 and CG2 can accommodate the preferences of Holists and Serialists.
The second implication is about why the EG (i.e., context version) cannot accommodate the preferences of Serialists. In essence, the vocabularies are presented in a relatively complex way in the EG, including the information about vocabularies, the context information, and relationships in the scenario, such as their locations, sizes, and colors. Thus, such complex and detailed information does not match with the preferences of Serialists, who tended to take a sequential or part-to-whole learning approach, resulting in some negative influences, such as cognitive overloading or the hesitation for the learning sequence. This reason might explain why the EG cannot support the preferences of Serialists.

Following the second one, the third implication is concerned with how the context version can be adapted to accommodate the preferences of Serialists. Due to the fact that Serialists follow a sequential or part-to-whole approach, the game-based learning environment should include some mechanisms that could guide them step by step. By doing so, they would not be put in a chaotic environment without any guidance. Accordingly, three mechanisms are proposed: index, structure, and story (Table 6). Regarding the index mechanism, students were offered a specific learning path via numbers or letters. For instance, using an alphabetical way to present information can offer students a linear sequential path. Regarding the structure mechanism, the context can be divided into a number of sub-contexts, and then can be presented to students one-by-one. For example, the living room can be divided into three sub-contexts of sofa, television, and bookcase. By doing so, Serialists would not receive a huge number of information at the same time. Regarding the story mechanism, specific objects can be highlighted by the camera movements and can be zoomed in/out. In other words, the story mechanism can use sequencing shoots, which can offer a clear learning path in a context.

Table 6. Three mechanisms suggested for Serialists

<table>
<thead>
<tr>
<th>Index</th>
<th>Structure</th>
<th>Story</th>
</tr>
</thead>
<tbody>
<tr>
<td>(using an alphabetical way to present information)</td>
<td>(dividing the context into a number of sub-contexts)</td>
<td>(presenting information by the camera movements)</td>
</tr>
</tbody>
</table>
Conclusions

Regarding the research question (i.e., How do Holists and Serialists react to different levels of scenarios in game-based learning systems?), the findings of this study revealed that (1) the learning performance of both Holists and Serialists was improved in the three different levels of game scenarios (2) Holists significantly made more improvement than Serialists in the context version of game scenario. It implied that the context version were appropriate to Holists for learning vocabularies whereas the context version were unsuitable to Serialists.

The contribution of this study includes two aspects: theory and application. In terms of the theory, this study deepens the understanding of the importance of cognitive styles in the development of game-based learning. The findings of this study indicated that cognitive style is a key factor that influences student learning in different levels of scenarios. Such findings can not only be used to improve the design of game-based learning systems in the future, but also are able to apply to develop scenario-based learning environments for various subject domains, e.g., language learning (Lin & Lan, 2015), science education (Chiang et al., 2014) and medical education (Pesare et al., 2016). In terms of the application, this study proposes three mechanisms, i.e., index, structure, and story, to support the needs of Serialists. Accordingly, designers can develop a personalized context for game-based learning systems with these three mechanisms. More specifically, we can incorporate the outcome of the SPQ into the development of personalized context, where these three mechanisms will be employed to provide additional support for Serialists.

However, this study has some limitations that should be further investigated in the future. First, this study was a pilot study with short-term treatment duration. However, there is also a need to examine long-term effects in the future. The other limitation is that the sample size of this work is relatively small. Thus, further investigation with large sample size is required. By doing so, we can develop a robust framework for the development of scenario-based learning based on the results of such future works and those of the study presented in this paper.

Acknowledgements

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan for their financial support under contract number (MOST-101-2511-S-155-004-MY2, MOST-103-2511-S-008-010-MY3).

References


Flow Experience and Educational Effectiveness of Teaching Informatics using AR

Stefanos Giasiranis1 and Loizos Sofos2

1Department of Primary Education-Postgraduate Studies Programme in Education Sciences-Education with the Use of New Technologies, University of the Aegean, Rhodes, Greece // 2Department of Primary Education, University of the Aegean, Rhodes, Greece // giasiranisst@gmail.com // lsofos@rhodes.aegean.gr

Corresponding author

(Submitted May 21, 2016; Revised October 25, 2016; Accepted October 31, 2016)

ABSTRACT

The purpose of this study was the investigation of the added value of technology of augmented reality (AR) in education and, particularly, whether this contributes to both student performance improvement, as well as the appearance of the psychological condition of Flow, which according to research, has had a positive effect on their performance when experienced during learning process. The research involved a total of 42 students in their second year of junior high school who were taught the module “Representation of the information on computers” using two different technologies, those of AR and the Web. Research data showed that both technologies contributed to student performance improvement and to the appearance of Flow to pupils, with apparently better results with the student group who utilized the technology of AR, though.

Keywords

Augmented reality, Flow, Interactive learning environments, Secondary education

Introduction

Only a few decades ago, few people had the privilege to have the availability of technology that could help them learn. Nowadays, technology in general and educational technology in particular has been rapidly evolving as well as being utilized in both formal and informal education. Computers, mobile phones, interactive whiteboards, simulations, virtual reality and Web 2.0 applications are just some technology examples, which have been effectively used by teachers and students in educational environments (Dror, 2008).

Nowadays, the new technology of AR has emerged in the field of education and up-to-date research shows that its use can have very positive learning outcomes. Such examples constitute research projects by Kerawalla, Luckin, Seljeflot and Woolard (2006), who searched the potential of AR in teaching the Earth-Sun interaction and day-night consecution, the EcoMobile programme (Kamarainen et al., 2013) concerning the use of the particular technology in environmental education and a large number of research games in open spaces, such as Outbreak at MIT, Environmental Detectives, Gray Anatomy etc. (Dunleavy & Dede, 2014).

In Greece, AR has been slightly used in education. The majority of applications concerned its use in open spaces of archaeological interest or inner museum and technology park spaces (Gialouri, 2011; Grigoraki, Politi & Tsolakos, 2013; Siampanopoulou, 2014; Sintoris, 2014). However, cases where AR is used in the classroom, such as the case of Dimitriou (2009) who created an AR application for the teaching of electrical circuits to high school students, almost do not exist.

Thus, the research on the use of AR in a classroom and the total absence of its applications for the subject of Information Technology is relatively small scale, at least in Greek reality. The present research was carried out in order to fill in both gaps, contributing to the further investigation of its pedagogical value. The answers to be given upon its completion, can highlight a different aspect of the use of AR in the educational process, encourage more researchers to explore its educational value, not only for the subject of IT in junior high school but also for other subjects and educational levels and, finally, inform teaching practitioners about the new technology and motivate them to start using it more often during their teaching sessions.

Review of relevant research projects

Students, coming in contact with the technology of AR for the first time, are impressed by the way virtual elements are incorporated into the environment they are located, and, as a result, they are motivated and actively participate in course activities (see Table 1).
During teaching sessions, they express their enthusiasm for what they do, collaborate on a great degree with one another in order to achieve their objective and, in a lot of cases, are absorbed in what they do in such a degree that they sense a modification of time or decreased reflex.

At the end of the teaching session, they have a positive attitude towards the technology used and claim to be eager to use it again. They consider it effective because it helped them learn and apt to help them learn more, although do not hide their satisfaction for what they have achieved by using it.

The accuracy of the students’ views seems to reflect on their learning outcomes. After the use of technology, they have better performance than before, they are able to observe objects, which, under normal circumstances, they are not able to, either because of their size (too big or too small) or because they are not visible in the environment, they retain their knowledge for longer periods.

<table>
<thead>
<tr>
<th>Study</th>
<th>Before teaching</th>
<th>During teaching</th>
<th>At the end of teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahn, and Choi, 2015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cai, Chiang and Wang, 2013</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Cai, Wang, and Chiang, 2014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cai, Wang, Gao, and Yu, 2012</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Chen, Liu, and Lu, 2013</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Di Serio, Ibáñez and Kloos, 2013</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Dunleavy, Dede, and Mitchell, 2009</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Dünser et al., 2012</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fleck, and Simon, 2013</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Freitas and Campos, 2008</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Ibáñez et al., 2014</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Juan, Toffetti, Abad and Cano, 2010</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Kamarainen et al., 2013</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Kerawalla et al., 2006</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Klopf, Perry, Squire, Jan, and</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Steinkuehler, 2005</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Liarokapis and Anderson, 2010</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Lin, Duh, Li, Wang and Tsai, 2013</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Liu, and Chu, 2010</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Nischelwitzer et al., 2007</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Pasaréti et al., 2011</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Salvador-Herranz et al., 2013</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Seo, Kim and Kim, 2006</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Shelton and Hedley, 2002</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Sin and Zaman, 2010</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Tarng, Ou, Yu, Liou, and Liou, 2015</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Wijers, Jonker, and Drijvers, 2010</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Wojciechowski, and Cellary, 2013</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Theoretical framework

The term AR refers to such technology which increases the sense of reality, allowing the coexistence of digital and factual information in the same environment (Azuma, 1997). The user is capable of not only simply seeing digital elements but also communicating and exchanging data, interacting with them.
Research in web environments (Liao, 2006; Shin, 2006; Webster, Trevino and Ryan, 1993), in games and in virtual reality environments (Faiaola, Newlon, Pfaff and Smyslova, 2013; Papastergiou, 2009) have showed that students’ learning outcomes can be enhanced if students experience the psychological condition of Flow during teaching. AR as a means which shares common features with virtual reality is expected to help students develop Flow.

The state of Flow can be described as the psychological situation of someone who is involved in a pleasant and enjoyable, for themselves, activity in the course of which they appear to be totally preoccupied in what they do. In order to be found in such a psychological situation, they have to meet two factors which play the most important role: (a) the perceived by them, difficulty of challenge they have to face, and (b) the perceived by them, skill to deal with this challenge. Therefore, even a low difficulty activity is able to induce Flow state when there is balance between these two factors. In the case of imbalance, a person can feel Anxiety when they consider that they have a lower degree of skills than those needed to complete the activity and Boredom when the opposite happens. The relation between these two factors has been represented on a model (Figure 1), where the psychological state of Flow constitutes a channel (Csikszentmihalyi, 1975).

Generally speaking, nine factors relate to the appearance of Flow (Jackson and Csikszentmihalyi, 1999): (1) challenge-skill balance, when both are at a high level and in balance with one another, (2) action-awareness merging, when everything occurs spontaneously and automatically, (3) clear goals, when the person knows what to do, (4) unambiguous feedback, when the person immediately knows whether they have achieved their goals, (5) concentration on task at hand, when the person is fully concentrated on and preoccupied with what they do, (6) sense of control, when the person feels they have their actions under control and can cope with anything which may occur (7) loss of self-consciousness, when the person loses their sense of self, (8) transformation of time, when the person feels that time has passed very quickly, or has lasted for centuries and (9) autotelic experience, when the person considers that the effort made was worth it.

![Figure 1. Initial model of flow (Csikszentmihalyi, 1975)](image)

A necessary condition to make the use of each and every form of technology effective in an educational framework is its proper teaching use (Sofos, 2011). What is important is not technology itself, but the way it is used to support learning (Bronack, 2011).

A teaching session is characterized successful by the degree of achievement of learning outcomes expected by the teacher at the end of each session. The existing objective difficulty in this kind of control is the way in which what the student has learned will be reliably tested, since the biggest part of his thought is not visible to others. To overcome this difficulty, the teaching practitioner resorts to search for clues that will certainly indicate knowledge acquisition. These clues become visible through an expected behaviour determined during lesson planning and described with the learning objectives and performance objectives. The learning objectives are more generally set in relation to the performance objectives. Therefore, there may be the case where a learning objective may be equivalent to a set of performance objectives. However, both describe an action or behaviour which can be observed and thus be controlled (Oosterhof, 2010; Rellos, 2006).

What must be ensured during objectives description is that a student’s performance constitutes a representative indicator of the skill being tested. What can help at this stage is the knowledge of the skills types as proposed by Bloom (1956) and formed the basis for two out of the three categories used by modern cognitive psychologists, that is the declarative and procedural knowledge (the third is problem solution) (Oosterhof, 2010).
Declarative knowledge corresponds to the first step of Bloom’s objectives taxonomy, Knowledge (Oosterhof, 2010). The purpose of learning happening here is the storage of information in the student’s memory and its recall and presentation later, almost in their original form. Procedural knowledge, on the other hand, corresponds to the remaining steps of Bloom’s taxonomy, Understanding, Implementation, Analysis, Synthesis and Evaluation. It is the form of knowledge to be acquired by a student in order to be able to complete an activity and often involves motor skills and cognitive strategies. To evaluate procedural knowledge it is useful to subdivide it into discriminations, concepts/notions and rules and follow a different technique for each one of them. Discriminations refer to students’ reaction to stimuli perceived by their senses and their evaluation is done by asking them to identify the stimulus which is different to the rest. Concepts/Notions refer to examples with particular characteristics which the students are again invited to locate. Finally, rules refer to the principles implementation and ask students to apply them to unknown examples (Oosterhof, 2010).

Purpose and research questions

The aim of this study was to investigate the contribution of AR technology to the improvement of student performance and the emergence of the psychological state of Flow through a teaching intervention to junior high school second-year students. These students would be taught the “Representation of the information inside a computer” module which is suggested in the curriculum using a digital implementation of AR. The results would be compared to the results of a second, equivalent, group of students who would be taught the same module using a different kind of technology, in particular the Web technology.

In order to achieve the goal of this research, the following research questions were posed:

- What kind of differences appear between the two groups after the teaching intervention, as far as their overall learning level and the individual categories of knowledge are concerned?
- Did students in each group experience the psychological state of Flow using their digital applications and which group appeared the strongest state?
- Did the groups show differences in each of the nine factors related to the psychological state of Flow and how big they were?

Method

The research was conducted in the Junior High School of Massari, a regional school on the island of Rhodes. The second-year class had a total of 42 students divided into two parties, initially equivalent to each other as shown by their performance in the positive subjects of the previous school year. The students of the first party (B1) were 20 while the second party (B2) had 22 students.

The first party constituted the control group, while the second, the experimental group. The students of both groups were taught “Unit 1-Digital World” of the school book for the subject of Informatics Technology in Junior High School, following a constructive approach and specifically a teaching scenario of Anchored Instruction. Anchored Instruction is based on the existence of an “anchor” which usually takes the form of a video. The video-anchor sets a problem and gives students the initial information in order to start solving it. The difference in the teaching approach is identified in the digital tool for the collection of extra information used by each group. The control group used the computer lab computers to collect information from the website http://diadiko.weebly.com in order to solve the problem set by the anchor, while the experimental group used tablets to collect information from an application of AR. Both the website and the application of AR were created for the research purposes with Weebly and Layar (basic version) tools, respectively, and their contents did not differ.

The AR application which was created recognized five different images. The first one, regarded general information about the Binary system, the next three regarded the codification of the text, the numbers and the images respectively and the last one, the process of decoding of the text, the numbers and the images. In order an image of those five to be recognized and the augmented information to be displayed on the screen of the tablet that a student was using, he/she had to enable both the wifi of the tablet and the Layar browser and then to focus with the camera of the tablet on the image. After the recognition the student could interact with the application by tapping with his/her finger the parts of the application in order to collect the necessary information.

The first research question referred to the investigation of the pedagogical value of the technology of AR compared to Web technology and their research data was collected through a quiz given to students beforehand.
and one week after the teaching intervention. The quiz was developed by the researcher for research purposes and numbered a total of 21 questions, nine of which related to declarative knowledge and the remaining 12 related to procedural knowledge, of which four questions referred to Concepts and 8 to Rules.

The other two research questions investigated the occurrence of the psychological state of Flow and the estimation of its intensity degree. The research data was collected using two different questionnaires developed by other researchers. Both were translated and adapted to the knowledge level of the students.

The first one (intermediate Flow questionnaire) was developed by Pearce, Ainley and Howard (2005) and its purpose was to assess more accurately the fluctuation of the Flow, which is more difficult to assess with one and only questionnaire given to students at the end of the research, especially in small-scale surveys such as this. This questionnaire was given to the students in two different teaching phases. It contained two questions (i.e., (a) How difficult have you found this activity? (b) How do you judge your skills during this activity?) of the 5 rank Likert scale and investigated the existence of balance between the difficulty of the activity completed by the students using their digital applications and their skills.

The second (final Flow questionnaire) given to students at the end of the research was developed by Jackson and Marsh (1996) had a Cronbach’s reliability indicator $a = 0.83$. It included a total of 36 questions of the 5 rank Likert-type scale (1 = strongly disagree, 2 = disagree, 3 = neither disagree nor agree, 4 = agree, 5 = strongly agree). These questions tried to seek the nine factors which the appearance of Flow is related to and every factor corresponded to four questions which were repeated, differently formulated, every nine questions.

In order to answer the research questions of the present study, the research data were analysed both descriptively using the Microsoft Excel 2010 programme and inductively using the statistical programme SPSS 19. More, specifically, as far as the inductive analysis is concerned, there was initially a regularity control of the variables through the Shapiro-Wilk test and then, for those variables presenting regularity, depending on the research question, what was chosen was either a parametric t-test of either dependent or independent samples. For the rest of the variables, we selected the corresponding non-parametric test, either the Wilcoxon one or the Mann-Whitney one.

**Results**

**Research question 1**

To determine the difference in the learning level of the two groups, their research data in the quiz after the teaching intervention was used. Initially, there was a variables’ regularity control through the Shapiro-Wilk test, since the sample ($n = 42$) was smaller than the limit of 50 people. Results pointed out that no variable showed regularity.

What followed was a variable control through the non-parametric Mann-Whitney test (Table 2), which showed that between the groups there was a statistically important difference on a significance level $p \leq .05$, hence heterogeneity between the two groups, only as far as the variable Procedural knowledge-concepts [$U(42) = 104.00, p < .05$] is concerned. In the other categories of questions and on the overall test performance, despite the fact that the experimental group had better results, the groups showed no significant difference between them.

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Experimental group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Declarative knowledge</td>
<td>3.90</td>
<td>2.511</td>
</tr>
<tr>
<td>Procedural knowledge-concepts</td>
<td>2.25</td>
<td>1.118</td>
</tr>
<tr>
<td>Procedural knowledge-rules</td>
<td>5.80</td>
<td>1.473</td>
</tr>
<tr>
<td>Procedural knowledge-total</td>
<td>8.05</td>
<td>2.328</td>
</tr>
<tr>
<td>Total score</td>
<td>11.95</td>
<td>4.478</td>
</tr>
</tbody>
</table>

*Note. $p \leq .05$.*

The two groups showed their smallest difference in the category Procedural knowledge-rules, with just 0.06 points in favour of the experimental group, whereas in the category Procedural knowledge-concepts as well as in the variable Total Procedural knowledge, presented their biggest difference with 1.07 and 1.13 points.
respectively, again in favour of the experimental group. Furthermore, they showed 0.92 points difference in favour of the experimental group in Declarative knowledge and 2.05 points difference in Total score, again in favour of the experimental group.

**Research question 2**

The answers of each student group in the intermediate Flow questionnaire were used to simulate, through the use of a table, the original Flow model (Figure 1) of Csikszentmihalyi (1975). Moreover, in order to assess the intensity of each situation, it was considered that the closer to a Flow state (diagonal) students find themselves, the smaller the degree of Anxiety or boredom they experience and respectively, the farther away from the Flow State they are, the greater the Anxiety they experience.

Two different tables for each group were created. The first concerned the psychological condition of students after the end of the first activity in which they used their digital applications, and the second concerned the psychological state of Flow after the end of the second similar activity.

At the end of the first activity, most students of the control group (Figure 2A) were in a state of Anxiety (n = 9, f = 45.0%), then to a state of Boredom (n = 8, f = 40.0%) and less to a state of Flow (n = 3, f = 15.0%). Of the students who were in the state of Anxiety, two seem to worry less than the others and were very close to the state of Flow, six were in a medium state of Anxiety and one was in a great state of it. Of the students who were in the state of Flow, two estimated that they were in a medium state of it and one in a small state of it. Finally, of students who were in the state of Boredom, four appreciated that they were slightly bored and very close to pass to the state of Flow, two in a medium state of Boredom and the other two at a large state of it.

At the end of the second activity (Figure 2B), there was an increase of students’ skills who were in the state of Anxiety, without, however, a change in their total number (n = 9, f = 45.0%). Six of them experienced a low degree of Anxiety and very close to the state of Flow, two of them experienced medium Anxiety and one felt great Anxiety. The number of students who found themselves in a state of Flow had increased (n = 4, f = 20.0%) and all of them felt a great degree of Flow. Finally, the state of Boredom was experienced by a student fewer than in the previous activity (n = 7, f = 35.0%). Of the aforementioned students, six felt a low degree of boredom and very close to the state of Flow, while only one student experienced Boredom of a medium degree.

For the experimental group, at the end of the first activity (Figure 3A), there seemed to be a balance between the number of students who were in the state of Anxiety and Boredom (n = 9, f = 40.91%), while the minority of students were as well in the state of Flow (n = 4, f = 18.18%). Of the students who were in state of Anxiety, five worried to a lower extent than others and were very close to the state of Flow, two worried a bit more and experienced Anxiety of a medium degree, while the other two worried even more and were in a great Anxiety State. Of the nine students who were in the state of Boredom, six experienced it at a lower degree and were very close to the state of Flow, one experienced it at a medium degree and the other two students felt a medium grade of Boredom. Finally, three out of the four students who were, according to their estimation, in the state of Flow, experienced it at a medium degree and one of them at a low degree.
At the end of the second activity (Figure 3B), there was a change in student estimation, but the balance between Anxiety and Boredom still maintained to a lower degree than before ($n = 8, f = 36.36$%), though. The number of students who estimated that they were in a state of balance between activity difficulty and their skill (Flow) had increased to six ($n = 6, f = 27.27$%) but were still fewer than the others. Students, who experienced a low degree of Boredom and very close to the state of Flow, had increased by one. Moreover, the number of students who believed that their skills had improved but they were still in a state of Boredom had increased as well. Finally, only one student remained at a medium degree of the state of Boredom, while two students, who were at a high degree in the state of Boredom towards the end of the first activity, had changed estimate. Similarly, the number of students who experienced a low degree of Anxiety and were very close to the state of Flow had increased by one, one student fewer than before was at a state of medium Anxiety, while one was still in a state of high Anxiety. Finally, of the students who were in the state of Flow, four estimated that they were in a state of medium Flow and two others at a state of high Flow.

Finally, the total score of each group was calculated according to the research data of the final Flow questionnaire. The scores of the control group students ranged from 83 to 140 ($M = 111.95, SD = 15.76$) and those of the experimental group students from 78 to 146 ($M = 123.27, SD = 16.84$).

**Research question 3**

The answers to the questions for each factor for each student were added and their average was calculated. The results showed that the experimental group had bigger averages in all factors and all of them ranged above the limit of 3.0, that is, the limit of neutrality according to the Likert scale of the questionnaire. On the other hand, the control group did not exceed the neutral threshold (3.0) in all factors and, therefore, students did not experience Flow in these factors (Table 3).

<table>
<thead>
<tr>
<th></th>
<th>Balance of challenges-skills</th>
<th>Action-awareness merging</th>
<th>Clear goals</th>
<th>Unambiguous feedback</th>
<th>Concentration on task at hand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Control</td>
<td>2.95</td>
<td>0.81</td>
<td>2.81</td>
<td>0.82</td>
<td>3.05</td>
</tr>
<tr>
<td>Experimental</td>
<td>3.44</td>
<td>0.63</td>
<td>3.02</td>
<td>0.70</td>
<td>3.57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sense of control</th>
<th>Loss of self-consciousness</th>
<th>Transformation of time</th>
<th>Autotelic experience</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Control</td>
<td>3.14</td>
<td>0.49</td>
<td>3.55</td>
<td>0.80</td>
</tr>
<tr>
<td>Experimental</td>
<td>3.28</td>
<td>0.72</td>
<td>3.67</td>
<td>0.75</td>
</tr>
</tbody>
</table>

To determine the existence or non-existence of a statistically significant difference for each factor between the two groups, the average of each factor was initially checked against the Shapiro-Wilk test on whether they fulfilled the regularity criterion, since the sample was lower than 50 ($n = 20/22$). The test demonstrated that all
factors showed regularity in both groups ($p > .05$), thus, the parametric $t$-test of independent samples was chosen, which showed that statistically significant difference between the two groups appears to be in the following factors: Balance of challenges-skills [$t(40) = -2.226, p < .05$], Clear goals [$t(40) = -2.330, p < .05$], Unambiguous feedback [$t(40) = -2.361, p < .05$], Transformation of time [$t(40) = -2.071, p < .05$] and Autotelic experience [$t(40) = -2.218, p < .05$].

Finally, to determine whether the statistically significant differences between the two groups were strong, the effect size indicator $d$ of Cohen (1988) was calculated, only for the particular factors. This indicator showed that the difference between the two groups was great, since, for all factors, it ranged between 0.5 and 0.8 (Balance of challenges-skills: 0.70, Clear goals: 0.74, Unambiguous feedback: 0.75, Transformation of time: 0.65, Autotelic experience: 0.70).

**Discussion**

**Research question 1**

Comparing the results of the two groups it seems that, even though both groups improved their results, the experimental group had better performance in all knowledge categories.

More specifically, the control group improved their average in Declarative knowledge questions by 3.15 points, the Procedural knowledge-concepts questions by 1.5, the Procedural knowledge-rules questions by 3.8, the total Procedural knowledge by 5.3 and the Total score by 8.45. Similarly, the experimental group improved their average in Declarative knowledge questions by 3.77 points, the Procedural knowledge-concepts questions by 2.46, the Procedural knowledge-rules questions by 4.41 the total Procedural knowledge by 6.87 and the Total score by 10.64.

**Research question 2**

At the end of the first activity during which the control group students used their digital application they seem to be dominated by Anxiety, then Boredom and finally Flow but at a lower degree. Most of them experience Anxiety at a medium degree and Boredom at a low degree.

At the end of the second activity, some changes in the psychological state of students are observed. Although the percentage of students experiencing Anxiety remains the same, the percentage of Flow increases, and, at the same time, the percentage of Boredom decreases. However, if we examine the intensity extent of each state, there emerges both a shift toward the state of Flow as well as an increase of the number and degree of students who clearly experience Flow.

Moreover, through the final questionnaire of Flow analysis, there seems that several of the control group students experienced Flow, not only during activities but in general as well, achieving a score which reached up to 140 points. The general average of the 11 students who experienced Flow, even though marginally, rated from 3.08 to 3.89, while only one student reached 4 (Strongly agree), with an average 3.89.

Similar results emerged for the experimental group as well, who experience, towards the end of the first activity, Anxiety and Boredom at the same percentage to the previous one and Flow at a lower degree. As far as student intensity in the states of Anxiety or Boredom is concerned, this is more of low degree whereas Flow mainly appears at medium intensity.

At the end of the second activity, Anxiety and Boredom continue dominating the group members, even though at low intensity, and the psychological state of Flow follows with an increased percentage. As far as the intensity of each state is concerned, low degrees of Anxiety and Boredom dominate as well as medium degree of Flow. However, now, two students experience high degree of Flow, an element which did not appear in the control group. Consequently, the number of students who are close to or already and clearly experience Flow increases, while, at the same time, the number of students who clearly experience Anxiety or Boredom decreases.

The score at the final Flow questionnaire of the experimental group students shows that they experienced as well a state of Flow, not only during activity but also generally, achieving a score which reached 146 points.
Research question 3

The analysis of the final Flow questionnaire showed a statistically significant difference between the two groups only in five out of nine factors relating to the psychological state of Flow. The averages of those five factors in which the two groups showed a statistically significant difference, were all greater for the experimental group. Moreover, Cohen’s (1988) indicator d showed that the difference between the two groups was great.

Therefore, we can conclude that the implementation of AR helped students learn what they had to do and when they had achieved their goals, they were concentrated more during activities experiencing balance between challenge and skills and in the end they felt that their effort was worth it, thus being more satisfied.

Conclusions

The results enlighten two different potentials of AR, which appear here at a greater extent than in Web technology. The first one is that it contributes, to a great extent, to students’ performance improvement and the second one is that it helps students experience the psychological state of Flow, which, in turn, helps them improve their performance.

A possible explanation of these results is that the augmented technology engenders impression and interest to students, which has as a result to motivate them more, to participate more actively and with more enthusiasm in course activities, to be more concentrated and comprehend better anything they are taught.

These results are consistent with the results of other relevant studies (Ahn and Choi, 2015; Cai et al., 2012; Cai et al., 2014; Chen et al., 2013; Dünser et al., 2012; Fleck and Simon, 2013; Ibáñez et al., 2014; Juan et al., 2010; Kamarainen et al., 2013; Lin et al., 2013; Liu and Chu, 2010; Nischelwitzer et al., 2007; Pasaréti et al., 2011; Salvador-Herranz, et al., 2013; Seo et al., 2006; Shelton and Hedley, 2002; Sin and Zaman, 2010; Tarng et al., 2015; Wijers et al., 2010). Although they cannot be used in general, it cannot be refused that they constitute clues of the pedagogical values of AR. Future research, which will surpass the present research restrictions (e.g., the small sample and duration), could certainly prove at a greater extent that these clues are, finally, valid.

References


Let’s Draw: Utilizing Interactive White Board to Support Kindergarten Children’s Visual Art Learning Practice

Pao-Nan Chou¹, Chi-Cheng Chang², and Mei-Yin Chen³

¹Department of Education, National University of Tainan, Tainan, Taiwan // ²Department of Technology Applications and Human Resource Development, National Taiwan Normal University, Taipei, Taiwan // ³Department of Kindergarten, Yong Kang Sheng Li Elementary School, Tainan, Taiwan // pnchou@gm2.nutn.edu.tw // samchang@ntnu.edu.tw // eternal5230@yahoo.com.tw

*Corresponding author

(Submitted June 7, 2016; Revised September 10, 2016; Accepted November 2, 2016)

ABSTRACT

Compared to other academic disciplines, interactive white board (IWB) research in early childhood education is still in its infancy. To add more knowledge base regarding the instructional effectiveness of IWB for young children, this study aimed to investigate educational phenomenon of using an IWB to teach visual art to kindergarten students. The study adopted a qualitative case study to fulfill the research purpose. The research participants were 25 children aged 5 to 6 years old from a public kindergarten in Taiwan. The qualitative study lasted for four months of one semester. Multiple qualitative data sources were collected from students, teachers, and parents. A qualitative analysis method identified seven major themes. The results indicate that the IWB is an effective learning tool for enhancing students’ learning interest and motivation during visual art instruction. The IWB may facilitate discussion about learning process, resulting in more diverse elements in children’s paper-based drawings. Schoolteachers and parents both witnessed the positive effect of IWB use on children’s drawing instruction. In particular, IWB integration strongly supports students with special learning needs. Based on IWB implementation experiences, several instructional implications for early childhood educators who are interested in IWB adoption in classrooms are proposed.

Keywords

Information technology in kindergarten, Visual art learning, Interactive white board use, Interactive learning environment, Drawing instruction

Introduction and problem statement

Problems of Information and Communication Technology (ICT) use in Taiwanese kindergartens

The Ministry of Education in Taiwan outlined the importance of ICT applications in classrooms for K-12 teachers. To foster the development of digital citizens, the Taiwanese government strongly encouraged teachers to embrace the concept of curriculum integration with ICT by holding several workshops on ICT applications in education (Ministry of Education, 2008). However, similar to the status and development in Western countries (e.g., Hsin, Li, & Tsai, 2014; Yelland, 2005), many early childhood educators in Taiwan remain skeptical about ICT use in kindergartens. Those who were unwilling to adopt ICT in the curriculum often expressed that ICT use might lead to impaired eye development, poor social interaction, and computer addiction (Chen & Huang, 2007). Chen and Huang (2007) further explored the factors influencing kindergarten teachers’ ICT use and found that teachers who resisted using ICT in the classroom often lacked professional knowledge regarding ICT use for curriculum reform and personal experience in utilizing ICT in their teaching practice. In addition, according to a recent survey report (Chen, 2011), from the preschool teachers’ perspectives, the largest barrier for ICT integration in kindergartens is administrational support.

Emphasis on traditional drawing instruction in Taiwanese kindergartens

Because young children in kindergartens have not received literacy training yet, they often draw symbolic images to express their ideas. According to Piaget’s (1956) constructive learning theory, children use symbols in drawings to construct their imaginary worlds. Kress and Van Leeuwen (2001) considered children’s visual artworks to be “image grammar” that communicates their conceptual knowledge to adults. Therefore, most kindergartens in Taiwan emphasize drawing instruction in the curriculum design (Shen, 2003). Several kindergarten teachers even examined the content of drawings as a benchmark of cognitive development (Ehrlén, 2009).
Drawing instruction is a subcategory of visual art education. Traditionally, kindergarten children’s drawings are created using pencils, pens, and crayons because schoolteachers emphasize on traditional teaching methods. For instance, according to learning resources listed in the Association of Taiwan Art Education (2015), most preschool teachers still employ traditional teaching methods without the need of technology tools in drawing instruction. Integrating ICT into visual art education is uncommon in Taiwanese kindergartens. This phenomenon echoed Terreni’s (2010) observations in worldwide early childhood settings, and also became a driving force in completing the current study.

Unpopularity of interactive white board use in Taiwanese kindergartens

An interactive white board (IWB) is a large touch screen display mounted on a wall. When used in classrooms, it is often combined with a digital projector that displays electronic signals from a desktop or laptop computer. Teachers or children can easily use electronic pens to annotate texts and move learning objects (e.g., images or animation) on the IWB (Preston & Mowbray, 2008). These interactive features situate young children in a joy-based learning environment (Morgan, 2010), where students’ learning interest and motivation are higher than they are in traditional learning environments (Terreni, 2010). Despite the claimed benefits of IWB use, according to Tsai’s (2013) survey report, the installation rate of IWBS in Taiwanese kindergartens was extremely low. Furthermore, based on the researchers’ long-term observation, a combination of IWB instruction and drawing instruction for young kids was also rarely seen in the kindergartens.

Theoretical discussion and research purpose

Foundations of IWB use in Kindergartens

According to a new position statement of the National Association for the Education of Young Children (2012), when used wisely, ICTs are effective learning tools that may support learning in young children. Similarly, the International Society for Technology in Education (2007) described the necessity of technological literacy standards for young children and emphasized ICT use in learning activities. Being as one of emerging ICT tools, therefore, IWB use in current kindergarten curriculum, particularly for visual art learning, perhaps presents a new instructional option in early childhood education.

According to the cognitive theory of multimedia learning (Clark & Mayer, 2011), IWB adoption in visual art learning might function as an effective learning tool that enables young children to systematically organize received information with prior knowledge. Furthermore, IWB use may serve as a scaffolding tool (Donohue, 2015; Jonassen, 1999) for helping children construct knowledge about drawings. IWB teaching practice, as a new instructional technology, perhaps can support and facilitate children’s paper-based drawing in traditional visual art learning.

IWB use with visual art learning in Kindergartens

Because the IWB is a relatively new ICT application, IWB research in early childhood education is still in its infancy, influencing the number of scholarly studies. In the relevant literature, most past research has focused on IWB integration with science-related subjects and reported positive findings. For example, Preston and Mowbray (2008) found that IWB adoption enhanced children’s science learning experiences. Linder (2012) indicated that IWB use was an effective learning tool for supporting children’s mathematics skills. Wong, Goh, and Osman (2013) revealed that IWB integration into the science curriculum improved the teaching and learning process. However, whether IWB integrated into visual art curriculum may yield similar outcomes remains further exploration in the current study.

Regarding a combination of IWB and visual art learning, after interviewing several UK kindergarten teachers about IWB use and observing young children behaviors during IWB integration in class, Morgan (2010) was the among the first to identify that IWB was a perfect tool to capture students’ visual ideas. Furthermore, Terreni (2011) integrated IWB use into a visual art curriculum in a New Zealand kindergarten, and reported that IWB use supported young children’s visual art learning experiences. However, because there have been few past studies on IWB use with visual art learning in early childhood education, the effect of IWB instruction on young children remains uncertain. Therefore, the implementation of the current study perhaps provides other new insights to IWB application in visual art curriculum.
Based on the previously discussed problem identification and theoretical background, one overarching goal of the study was to use a qualitative case study method to investigate the educational phenomenon of integrating IWB into visual art instruction in a public kindergarten in Taiwan. Specially, under the goal, the research objectives were:

- What were young kids’ responses to IWB application in drawing instruction?
- Compared to the previous semester, what kinds of new elements might appear in students’ drawings after IWB use?
- What instructional scenarios (or problems) might happen during IWB implementation?
- What were teachers’ and parents’ perceptions of IWB use in the visual art curriculum?

Research method

Research design

The current study involved long-term observation in one educational setting. A qualitative approach was the most appropriate research method for fulfilling the research purpose (Hatch, 2002). Of the varied qualitative paradigms, this study adopted a case study methodology to collect the required data because case study research enables researchers to fully concentrate on one specific case in a school (Creswell, 2007). According to Yin (2003), a case study can have a simple single-case or complex multi-case design depending on the unit of analysis. In this study, the main focus was how IWB integration affected children’s learning process and drawing learning whereby the single-case design fits this research’s scenario.

Research participant

The participants in the study were 25 children aged 5 to 6 years old. These children were students at a public kindergarten in a metropolitan city in Taiwan. One school teacher and two childcare workers who assisted the schoolteacher were assigned to the class. The school was located in a downtown area where the students’ parental backgrounds were more diverse than those in countryside areas. Before the implementation of the study, research consent forms were obtained from the students’ parents. To confirm with ethical standards, the names of the students and settings were assigned codes.

The schoolteacher had six years of experience of working in kindergarten where she often taught visual art classes. She obtained her bachelor’s degree of early childhood education and master’s degree of curriculum and instruction. Prior to the study, the schoolteacher had already received professional training regarding IWB use in class.

Prior to the study, the ICT infrastructure in the classroom contained a desktop computer and a digital projector. The students’ ICT exposure was limited to PowerPoint instruction models, where teachers used to introduce health issues and promote school policies to students. No learning opportunities regarding ICT manipulation in the school were provided to the students. Because the IWB was newly introduced in the kindergarten, the students had no previous experiences with using an IWB.

Curriculum design

To enable the IWB to be successfully integrated into the existing visual art curriculum, the researchers and the schoolteacher collaboratively designed eight learning units about drawing practice by modifying the official curriculum current used in the kindergarten. Two learning units were taught each month. Each learning unit lasted 2.5 hours. Prior to the study, the teacher introduced the IWB into the classroom, enabling the students to understand the technological functions embedded in the IWB. Table 1 lists the curriculum design of the study.

The curriculum design listed in Table 1 was coordinated with other learning activities. For example, before the start of Unit 1, the schoolteacher took the students to stroll around the kindergarten neighborhood. The activity enabled the students to observe the neighborhood so that they could have a reference to draw for their visual art class. Furthermore, the schoolteacher chose specific picture books related to the unit titles. This storytelling activity enabled the students to learn how to organize what they observed in their drawings. During the learning unit, the schoolteacher employed the IWB to demonstrate basic drawing knowledge and skills. Several students were also encouraged to draw their imaginary objects on the IWB. After 1.5 hours of IWB use, the students used
crayons to complete their artworks on a work sheet and share their stories with classmates. Upon completion of drawing instruction, homework (paper-based drawing) related to the learning unit was assigned to the students. Students were asked to practice their visual art learning on a piece of paper.

<table>
<thead>
<tr>
<th>Unit title</th>
<th>Description (Learning goal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is neighborhood?</td>
<td>Children observe something in the neighborhood</td>
</tr>
<tr>
<td>Symbol</td>
<td>Children understand shop symbols</td>
</tr>
<tr>
<td>Billboard</td>
<td>Children discover interesting billboard</td>
</tr>
<tr>
<td>Where are you going?</td>
<td>Children understand several routes in the neighborhood</td>
</tr>
<tr>
<td>Little receptionist</td>
<td>Children take a role of a receptionist</td>
</tr>
<tr>
<td>Little hair designer</td>
<td>Children take a role of a hair designer</td>
</tr>
<tr>
<td>Little cloth designer</td>
<td>Children take a role of cloth designer for Halloween</td>
</tr>
<tr>
<td>Shopping fun</td>
<td>Children buy something in a supermarket</td>
</tr>
</tbody>
</table>

*Note.* Learning units were coordinated with other two learning activities (field trips and storytelling with picture books)

**Interactive White Board in the classroom**

In the study, the IWB served as a scaffold for supporting teaching and learning in the class (Donohue, 2015; Jonassen, 1999) rather than replacing traditional teaching methods. The research assumption was that IWB use in the classroom might stimulate the students’ learning interest and inspire them to draw visual objects. When integrating the IWB into teaching practice, the schoolteacher employed IWB software to prepare learning materials, which contained several animations, static images, and Internet learning resources. Overall, IWB (a commercial tool in the market) establishment in the classroom provided a digital drawing function for the students. Figure 1 shows a student who drew objects on the IWB.

![Figure 1. Student drawing objects on the IWB](image)

**Data collection**

Yin (2003) suggested that a case study should involve collecting data from multiple sources. This study adopted seven data sources:

- Recorded instruction videos: During each learning unit, a digital camera recorded all learning scenarios. The researcher used these data to examine the instructional process of each learning unit.
- Teacher’s teaching notes: The schoolteacher wrote a self-reflection in her teaching notes after the completion of each learning unit. From a teaching perspective, the notes presented what the teacher saw in the class.
- Learning materials: All learning materials presented on the IWB were collected. These data were used to ensure consistency between the instruction videos and teaching notes.
• Peer feedback: Two schoolteachers in other classes were invited to observe the learning units by providing written peer feedback. These data offer an objective viewpoint on students’ learning process.
• Students’ work sheets and homework: Students’ in-class and assigned homework drawings were collected after the completion of the study. These drawings were compared with the students’ drawings in the previous semester (without IWB intervention).
• Parent survey: After all learning units were finished, a questionnaire was sent to the students’ parents. The survey consisted of five open-ended questions on the parents’ perceptions of visual art learning. The questions are listed in Table 2.
• Interviews with parents and students: After each learning unit, the schoolteacher conducted an informal interview (no semi-structured guidelines) with several parents and students. The style of the informal interview was a conversation focusing on drawing practice between the schoolteacher and different parents (without children) and students (without parents), which created a low-pressure environment. The informal interview was conducted within 15 minutes. The teacher summarized the interview results in interview notes.

Table 2. Open-Ended Questions in the Survey

<table>
<thead>
<tr>
<th>Item</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>How often did your kid draw at home?</td>
</tr>
<tr>
<td>Q2</td>
<td>How creative was your child’s drawing?</td>
</tr>
<tr>
<td>Q3</td>
<td>How did you perceive the role of IWB instruction in visual art learning?</td>
</tr>
<tr>
<td>Q4</td>
<td>After examining your child’s working sheets, how did you perceive his (her) visual art learning?</td>
</tr>
<tr>
<td>Q5</td>
<td>Any feedback regarding IWB adoption in the classroom?</td>
</tr>
</tbody>
</table>

Note. *The creative elements were based on parents’ judgment compared to the drawings created in the previous semester.

Data trustworthiness, coding and analysis

The study adopted a data triangulation method proposed by Patton (2002) to ensure qualitative data trustworthiness. Figure 2 illustrates the data triangulation method. Furthermore, the study employed a coding strategy to ensure information privacy. Table 3 summarizes the coding principle for each data source.

![Figure 2. Data triangulation method](image)

Because Teaching Note, Peer Feedback, Parent Survey, and Interview with Parents and Children were qualitative-based text information, the researchers recruited three graduate students majoring in education to transcribe original paper-based data and assigned coding name to each category of documents. Totally, more than 300 copies of A4-size documents (transcripts) were retrieved. The Recorded Video, Learning Material and Work Sheet and Homework were used for corroborating the transcripts.

Table 3. Coding strategy

<table>
<thead>
<tr>
<th>Data source</th>
<th>Coding name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recorded Video</td>
<td>RV</td>
</tr>
<tr>
<td>Teaching Note</td>
<td>TN</td>
</tr>
<tr>
<td>Learning Material</td>
<td>LM</td>
</tr>
<tr>
<td>Peer Feedback</td>
<td>PF</td>
</tr>
<tr>
<td>Work Sheet and Homework</td>
<td>WS and WH</td>
</tr>
<tr>
<td>Parent Survey</td>
<td>PS</td>
</tr>
<tr>
<td>Interview with Parents and Children</td>
<td>IP and IC</td>
</tr>
</tbody>
</table>
After all data sources were obtained, the researchers employed Moustakas’s (1994) four-stage analysis method to interpret qualitative transcripts. First, crucial phrases were identified. Second, similar phrases were grouped to form meanings. Third, the formulated meanings created several themes. Finally, a detailed text description with representative quotations for each theme was provided. During the process of data analysis, three researchers collaboratively analyzed the results from the four-stage analysis and discussed the data triangulation among seven data sources. Overall, seven themes related to research content were selected for further discussion.

**Research results**

Seven major qualitative themes with representative supporting data are presented in this section.

**Theme 1: Joyful learning environment activated learning motivation**

When the IWB was introduced into the classroom, the students showed their excitement toward drawing instruction. In each recorded video clip (from RV-Unit1 to RV-Unit8), excited sounds and expressions from students were captured during class. For example, “Wow! IWB again” and “Yeah! I love IWB” were common phrases that the students used to describe their feelings of excitement. In the teacher’s teaching notes, the teacher also observed similar situations. The teacher wrote one note: “Sounds of excitement filled the class. Each child was eager to see what the teacher would introduce by using the IWB” (TN-Unit1). Such joyful learning experiences might linger for a long period of time. Several interviewed parents recalled that their children would actively share how they used the IWB during class at home. For instance, one mother said, “My son often brought the IWB into conversation and told me that the IWB was very interesting stuff. He drew something on it” (IP-S6).

Once a joyful atmosphere was created in the learning environment, students’ learning motivation and interest were aroused (observation from RV-Unit1 to RV-Unit8). Compared with a traditional teaching method, the students focused more on learning activities designed by the teacher. IWB integration into drawing instruction enabled the students to experience new learning opportunities. The IWB acted as a learning hub that focused each student’s attention. The most intriguing finding was that some students who performed inactively in regular classes became interested after IWB integration. Changing attitudes enabled those students to actively respond to the teacher’s questions. Figure 3 shows the student’s active involvement in IWB instruction.

**Theme 2: IWB integration arousing learning interests**

When the IWB served as a learning tool in the classroom, the students always showed interests in using the IWB. However, because only one IWB was installed in the class, learning opportunities were limited for all students. Several video clips demonstrated that to attain learning opportunities to practice drawing on the IWB, the students were willing to stick to instructional rules established by the teacher. One feedback comment from a peer teacher stated, “The kids quietly raised their hands. No other noises were made. They wanted the teacher to call their names” (PF-Unit3). While selected students drew on the IWB, other students not only watched their drawing demonstration but also provided several drawing suggestions. One teaching note stated, “When the kids
drew their works in front of the IWB, other children enthusiastically discussed what their peers drew” (TN-Unit2).

During the IWB intervention, the students had to follow one rule: peers’ drawings cannot be criticized. After volunteers drew on the IWB, learning discussion was open to all students. First, the teacher praised the volunteers’ active participation and encouraged them to share the meaning of their drawings on the IWB. Second, other students discussed the volunteer’s drawings. Finally, other students who had drawn on the IWB expressed their viewpoints. Overall, from each recorded instruction video (from RV1 to RV8), students were engaged in a lively learning discussion. Figure 4 shows a student who volunteered to draw on the IWB.

**Figure 4. One volunteer for IWB drawing (Screenshot from RV-Unit3)**

**Theme 3: New art learning process**

The students were attracted to the IWB and thus students attempted to obtain the teacher’s permission to use it. The content of the video clips often indicated that specific students actively expressed their learning desires. To ensure that all students to have an opportunity to demonstrate their drawings on the IWB, the schoolteacher often self-reflected her choices. One teacher’s note also stated, “While some kids wanted to show something on the IWB, other students quietly listened to my lesson. I should remind myself to pick inactive students” (TN-Unit2).

To enable the students to understand the technological functions of the IWB, the teacher introduced the IWB into the classrooms several times prior to the study. However, during the IWB learning intervention, some students still showed awkward behaviors when standing in front of the IWB. For example, one peer feedback comment stated, “Even though the kids already learned to operate the IWB, some still hesitated when it was their turn to show something” (PF-Unit2). The teacher provided those students with learning scaffolding and helped them become familiar with IWB operation and confident to draw pictures.

Several interviewed students reported that the IWB was an easy-to-learn and fun tool. Compared with traditional teaching, they expressed enthusiastic feelings regarding using the IWB for drawing. For example, one boy said, “I do not like to use a plastic eraser to remove something on paper, but the IWB provides an easy function for erasing objects” (IC-S9). One girl stated, “I can easily choose any size of a pencil for drawing pictures. It is difficult to achieve this purpose for traditional drawing” (IC-S14). In addition, a few students extremely enjoyed IWB drawing practice. They asked the teacher if they could use the IWB to complete drawings for work sheets and assignments.

**Theme 4: IWB integration strongly supporting students who need more learning attention**

Prior to the study, three participants were identified as special care students who need more learning attention because several kindergarten teachers observed their inactive learning behaviors in classes. Table 4 lists the profiles of those students, who are denoted using codes. However, against expectations, during the IWB learning intervention, those three students performed better than they did during activities in traditional teaching environments. In general, the IWB aroused their learning interest to engage more in learning activities.
Table 4. Profiles of students who need more learning attention

<table>
<thead>
<tr>
<th>Pseudo name</th>
<th>Learning behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Shyness and few interaction with classmates</td>
</tr>
<tr>
<td>S2</td>
<td>Low learning confidence</td>
</tr>
<tr>
<td>S3</td>
<td>Low Learning achievement</td>
</tr>
</tbody>
</table>

Note. *These behaviors were not identified illnesses.

S1 demonstrated shyness in all classes. She seldom talked with other children in the classroom. After 3 weeks of the IWB intervention, S1 showed her interest in IWB operation. The teacher recorded this exciting moment in one note that stated, “I have taught this class for one and half years. This was the first time I saw S1 raise her hand” (TN-Uni4). Although S1 spoke quietly in front of the classroom, she shared ideas about her drawing with the other students. When interviewed about her changed attitude, S1 responded that she had already practiced a similar learning activity several times at home.

S2 often showed low confidence during learning activities. When the teacher asked if he had any ideas about drawings, he always said “No.” During drawing creation, S2 often sat quietly and gazed at his white work sheet for a long period of time. After a few weeks of the IWB intervention, S2 became more interested in classmates’ drawings. According to the teacher’s observation in one note, “S2 fully paid attention to other children’s drawings shown on the IWB” (TN-Uni5). When provided a copy of a work sheet, S2 quickly completed his drawing. When asked about his feelings about drawing, he responded that the IWB learning activity inspired his thinking and gave him more ideas.

S3 was active in sports activities outside the classroom. When learning occurred in the classroom, S3 often showed restless behaviors and did not pay attention to the teacher. At the beginning of the IWB intervention, S3 became interested in the new technology. When interviewed about his learning attitudes, he responded that he had experience playing with new technology at home because his parents bought a lot of technology products, such as a tablet computer, for him. Therefore, the teacher allowed him to play the role of “helper.” S3 demonstrated his excitement to assist other students in becoming familiar with IWB operation during the study.

Theme 5: Diverse ideas expressed through drawings

After the completion of the study, the teacher and other peer teachers compared the students’ paper drawings during the IWB intervention with those drawn without the IWB intervention. Regardless of drawing quality, on average, more diverse elements appeared in the students’ drawings with the IWB intervention. The teacher used peer imitation to describe this finding. One teaching note stated, “The IWB learning activities probably enabled the students to incorporate other students’ ideas in their drawings. Students use similar concepts to create their personal style drawings” (TN-Uni6). Other teachers perceived that a joyful learning environment played an important role. One peer feedback comment stated, “During IWB instruction, the learning environment inspired the students to think. More ideas popped up when the children demonstrated their works on the IWB” (PF-Uni4).

![Before](image1.png) ![After](image2.png)

Figure 5. One child created more colorful elements after the IWB intervention

When asked about the students’ drawing habits at home, several parents stated that the IWB learning activities facilitated the drawing development process, enabling the students to introduce diverse elements into their visual works. One mother responded in the survey, “I observed that my son began to try different visual elements to
express his drawings. The drawings did not look very boring” (PS-S11). Another mother said, “Usually, my kid’s drawings only contained one or two colors, but his drawing style became more colorful” (IP-S13). Although the IWB strengthened the students’ drawing content, many parents reported that their children’s drawing frequency at home did not substantially increase. Figure 5 shows one representative student’s (WS-S9) drawings before and after the IWB intervention.

**Theme 6: IWB integration as a focus in the kindergarten**

In this study, eight learning units designed in the visual art class were combined with other learning activities. Drawing instruction practice became a showcase in which the students demonstrated knowledge they received in the other two learning activities. Such theme-based learning enabled the students to focus on learning activities. One teaching note stated, “The kids knew that other activities were their idea inputs. They paid attention to the learning content. They did not want to miss something important” (TN-Unit8). Teachers of other classes in charge of learning activities expressed similar observations. One peer feedback comment stated, “The kids seriously observed things and listened to picture stories” (PF-Unit7).

Because IWB adoption occurred only in one classroom, teachers of other class in the kindergarten showed their interest in IWB integration into drawing instruction. Instruction with an IWB became a popular topic for discussion in the school. High levels of curiosity caused the teacher’s colleagues to observe the IWB’s practical impact on learning in children. The teacher recalled this phenomenon in one teaching note, which stated, “Everyone was curious about new things. My colleagues wanted to know if the IWB affected student learning” (TN-Unit6). Once colleagues witnessed the students’ learning during drawing instruction using the IWB, their skepticism regarding ICT use disappeared. They began to consider the possibility of technology integration into the existing curriculum.

**Theme 7: New problems during IWB instruction**

Although the new instruction method created a positive learning environment, some new problems occurred during the IWB intervention. First, the students’ emotions fluctuated throughout the school day. After the completion of a 2.5-hr learning unit, the students were ready to take other classes. However, the students’ joyful learning feelings suddenly disappeared when they faced a traditional teaching method. According to one teaching note, the teacher wrote, “My colleagues told me that the children obviously showed a sad feeling after the visual art class. It’s difficult to motivate them again” (TN-Unit8). Other teachers in the kindergarten even called those students “deflated balloons.”

![Figure 6. Roadmap of the study from analysis to evaluation](image-url)
Second, the IWB platform in the classroom was installed at one specific height. From a human and computer interaction perspective, the height enabled the teacher to easily access objects on the IWB. However, the fixed height was not suitable for children aged 5 to 6 years, causing several students to have difficulties in operating the IWB. Therefore, the teacher should have prepared a standing platform for shorter children. The teacher self-reflected in one teaching note, stating “If the IWB could be vertically adjusted to a certain height, that would be perfect for all kids” (TN-Unit3).

Last, compared with a traditional teaching method, a curriculum that involves computer use and technology tools requires more learning resources. Learning materials contained several multimedia elements (from LM-Unit1 to LM-Unit8). However, such a learning atmosphere created a teaching burden for the teacher, who had to spend a lot of time preparing learning materials by using computer software. Furthermore, to avoid spontaneous technical problems in the classroom, the teacher should demonstrate materials on the IWB and be familiar with tool manipulation before the beginning of the learning unit. One teaching note stated, “Extra effort should be put into IWB teaching preparation. Maybe we can create an IWB team in the future so that learning materials can be shared” (TN-Unit4).

Figure 6 is a roadmap that provides a scientific overview from analysis (problem identification) to evaluation (seven qualitative themes).

**Discussion**

Over the course of the eight learning units implemented in this study, the students were interested in IWB instruction, which directly influenced their learning motivation. This finding was consistent with past research that integrated the IWB into science learning in kindergartens (Linder, 2012; Preston & Mowbray, 2008; Wong et al., 2013). Regarding visual art learning, the finding also supported a previous study that indicated that the IWB might strongly arouse children’s learning interest and motivate students to participate in IWB-related learning activities (Terreni, 2011). According to Keller's (1983) motivation theory, stimulating students’ curiosity is a critical step in instructional design. In the study, therefore, the IWB served as a novel learning tool for increasing the students’ curiosity during the IWB intervention.

In the current study, the joyful learning environment created by IWB instruction obviously facilitated peer discussion and interaction. While volunteers shared their drawings on the IWB, other classmates enthusiastically discussed objects in the drawings and responded to the teacher’s queries. Furthermore, after school, the majority of students were eager to share IWB learning scenarios that occurred in their classrooms with their parents, which in turn facilitated social interaction at home. Therefore, the findings in the study echoed the analytical report by Hsin et al. (2014), which stated that ICT positively benefited children’s social development when used wisely.

In the study, the IWB enabled students with varied drawing capabilities to construct their individual-based drawings by socially interacting with peers and the teacher in a highly motivated manner. Under such a learning environment, the IWB use in the classroom indeed served as a useful scaffolding tool (Jonassen, 1999; Donohue, 2015) to support learning interaction. In addition, although the quality of the students’ drawings did not improve markedly, more diverse and colorful elements appeared in the drawings compared to visual art works in the previous semester. This finding can be attributed to social learning (Bandura, 1977) in children’s discussions. The students perhaps observed other styles of visual art and incorporated new elements into their works, which became unique drawings with personal styles.

According to a report from the United Nations Educational, Scientific, and Cultural Organization (UNESCO, 2006), ICT use in schools may support students with special learning needs. For example, ICT can “unlock hidden potential for those with communication difficulties … [and] enable students to demonstrate achievement in ways which might not be possible with traditional methods” (p. 30). In this study, IWB integration in the visual art curriculum played a critical role in supporting three students who needed more learning attention. Adapting the IWB in the classroom enabled one student with an introverted personality to overcome communication problems, one student with low learning confidence to develop more diverse ideas, and one student with low learning achievement to demonstrate potential learning abilities not shown in a traditional teaching environment. This finding supported a finding from Terreni’s (2011) study, which showed that IWB use benefited students with special learning needs in a visual art class.
User experiences reflected the value of technology adoption (Carr-Chellman, 2006). In this study, an IWB mounted on the wall seemed to be a large tablet computer screen on which the students could enjoy working on their drawings. Easy-to-use features in the IWB enabled the students to have new art learning experiences. Some students even perceived the possibility of IWB drawing replacing traditional drawing methods. Although unfamiliar feelings at the initial implementation stage were reported by some students, the user-friendly interface of the IWB still triggered students’ learning desire for drawing practice. Therefore, from a practitioner’s perspective, IWB adoption in the classroom fitted well with the visual art curriculum. In accordance with Morgan’s (2010) observation in UK schools, the IWB was perfectly used to capture students’ visual ideas.

Most previous studies tended to focus only on students’ learning performances. The role of parents was always a missing component of ICT integration in classrooms. In this study, parents’ perceptions were obtained to construct a multi-faceted perspective of ICT adoption (Ravasco et al., 2014). According to the analysis of qualitative themes, several parents identified their children’s changed drawing styles. More diverse and colorful elements appeared in the drawings. If IWB instruction is implemented during a school day, children might be eager to share classroom learning stories with their parents after school. Although IWB practice in school might not increase children’s drawing frequency at home, most parents appreciated the positive effect of the IWB on children.

From a pedagogical perspective, the IWB was an effective tool for facilitating the teaching process (Wong et al., 2013). However, because the IWB intervention was implemented only in the visual art related curriculum, the visual art teacher was fully responsible for technical preparation of multimedia learning materials. The extra effort required increasing the teaching load for the teacher. This finding was consistent with prior research that indicated that IWB integration in the classroom required an investment of time and effort for the instructors (Smith et al., 2005). Furthermore, when IWB instruction yielded learning benefits for the students, other teachers in the kindergarten began to consider the possibility of technology adoption in their classes. Thus, IWB instruction elicited a relative learning advantage that prompted teachers of other classes to accept ICT adoption (Rogers, 2003).

Conclusion

Response to research goal

The overarching purpose of this study was to explore the educational phenomenon of using an IWB to teach visual art in kindergarten. The analysis of qualitative themes revealed that the IWB effectively aroused the students’ learning motivation and facilitated the students’ learning discussion during visual art instruction, resulting in more diverse elements in their paper-based drawings. In particular, IWB integration unlocked latent capabilities for students who need more learning attention. Furthermore, the students enjoyed using technological interfaces in the IWB to demonstrate their drawing. Schoolteachers and parents both perceived that the IWB had learning benefits on the students’ drawing learning.

Research limitations and recommendation for future work

Because of the unique nature of a qualitative case study, several research limitations existed in the study. First, research design in the study emphasized on qualitative scenarios rather than an experimental comparison. Future studies may investigate the comparison between IWB integration and traditional instruction through a quasi-experiment. Second, the duration of IWB instruction in the study was 4 months. Future studies can extend the intervention scope to examine whether children still enjoy IWB instruction over a long period of time. Third, the learning content in the study was drawing practice. Future studies can introduce the IWB into other types of visual art learning, such as clay creation. Different learning responses to IWB instruction may be obtained. Forth, drawing instruction in the study was coordinated with other learning activities. Future studies can verify the effect of sole IWB instruction on children’s learning. Fifth, the tablet computers such as I-Pad provide a similar function for digital drawing. Future studies may evaluate if different tools may yield similar results. Last, the limited opportunity for using IWB and the novel effect of IWB may lead to the joy feeling of the students in class. Future studies may explore those two factors during IWB integration.
Instructional implications

IWB integration into the kindergarten curriculum is an innovative pedagogy. Based on implementation experiences identified in the study, some instructional implications are proposed for early childhood educators who are interested in IWB instruction. First, IWB adoption in classrooms requires professional training. Inadequate technical support may influence the usage intention of teachers. Second, to avoid additional teaching problems, a suitable height for installing the IWB for children should be established. Third, because most children enjoy using IWB to create drawings, the instructor may allow students to submit their digital works as assignments rather than paper-based drawings. Finally, during IWB integration, several students may attempt to obtain permission to use IWB. Each student should be given an equal opportunity for technology use.

References


Heuristics and Web Skills Acquisition in Open Learning Environments

Daniel Domínguez Figaredo
National Distance Education University (UNED), Faculty of Education, Department of Educational Theory and Social Pedagogy, Madrid, Spain // ddominguez@edu.uned.es

(Submitted June 10, 2016; Revised October 11, 2016; Accepted November 11, 2016)

ABSTRACT

Web literacy refers to the skills and competencies people need in order to function in societies connected through the Internet. Many of the frameworks for understanding the components of web literacy are limited in value because they rely on conceptual definitions. They do not take into consideration the social practices governing the use and writing on the web. Nor do these frameworks take into account the open and participative nature of the Internet. With the aim of moving beyond this theoreticist vision, we present an analysis of the relationship between the social practices of a group of university students in open learning environments and the acquisition of web skills. We proposed an alternative approach that is rooted in an understanding of social practices. In order to “operationalize” and facilitate an study of web skills, we relied on a specific type of analysis that allowed us to observe the consistency between the practices observed and the behavior reflected in the heuristic framework of web skills. The main elements of this alternative framework are explained, as is the link between the social practices of the students and the skills acquired. We also discuss other contributions to the field of Web Literacy and to the even larger field of Digital Literacy.

Keywords

Web skills, Web literacy, Digital literacy, Open learning, Non-formal learning, Heuristic approach

Introduction

The way digital skills are taught is a subject that has been studied since the late 1990s, when analog devices were beginning to be replaced by digital ones as a first-tier socialization factor. Since then, there has been a wide-ranging discussion on how the need to acquire web skills has had an impact on how people are trained to carry out ordinary daily tasks. Many studies have shown that there are substantial educational and psychological benefits to incorporating digital social media into education and work (Coward et al., 2014; Gasser et al., 2012; Junco, 2015). In addition, it is beneficial for professional careers. Therefore, it is necessary to provide citizens with the basic skills required to function in the information society. This is known as Digital Literacy Instruction. This concept appeared because of the accelerated changes that were taking place as a result of massive use of software and hardware connected via the Internet, and the fact that this new system was inaccessible to many people (Lankshear & Knobel, 2007).

The Web is currently the best forum for becoming digitally literate. This technology is part of the majority of mediation devices. Social environments can be created using the Web. This makes it an appropriate tool for facilitating all types of human activity and interaction. The Web is an emerging technology; therefore the approaches on how to teach the skills people need to use it have evolved over time (Belshaw, 2014; Hall & Tiropantis, 2012).

In the field of education, the development of Web literacy instruction has been greatly influenced by digital, media, and information literacy instruction. As with these other kinds of literacy instruction, research about Web literacy revolves around operationally identifying and classifying the skills that people need to acquire in order to be regarded as “literate.” One key focus of these studies is to clearly map the skills related to handling browser software and programming languages. A second key focus is determining what distinguishes these skills from one another (European Committee for Standardisation, 2014; Dore, Geraghty, & O’Riordan, 2015). The Internet is different from other digital media and tools. As a consequence, the skills and competencies required and the mechanisms by which they are acquired, may be very different from other types of literacies (Ahtikari & Eronen, 2004; Lankshear & Knobel, 2013).

The existing approaches to studying these issues are either not specific enough, or ignore the participative nature of the Web. Here we introduce a new approach that focuses specifically on social practices in web environments and their relation to the development of digital skills. The reference framework used in this study is Web Literacy Map. This method for grouping skills was designed by Mozilla Foundation in conjunction with stakeholders with formal and informal education, as well as industry (Belshaw & Hilliger, 2015). The Web Literacy Map is used as heuristic model that reflects the skills required to read, write, and participate effectively.
on the Web. The Web Literacy Map includes the skills that Mozilla users consider important for reading, writing, and participating on the Web. At the time of writing, the latest iteration was v1.1.

Web literacy and the skills and competencies needed to use the Internet

Web literacy instruction is a subset of the field of digital literacy instruction that shares some characteristics with other types of literacy instruction, such as media literacy, computational/algorithmic thinking, and computer science (Belshaw, 2013). From an educational standpoint, web literacy can be analyzed in a practical manner by dealing with the skills that a person needs in order to be considered “literate.” However, most studies of digital skills are conceptual analyses of the types of knowledge to acquire and the skills themselves. That is, they focus on their definition and operationalization (Bawden, 2008; Beetham, McGill & Littlejohn, 2009; Ala-Mutka, 2011; Ferrari, 2012). The many types of web participation and the range of possible behaviors possible are not taken into account. These should be a fundamental part of the analysis. As a reaction to this deficiency, it has become more common to study these skills directly by observing the practices of people on the Internet. In these new types of studies, an analysis of people’s interactions and what motivates them to participate in online forums is favored. The objective is to understand the skills that are developed and that are necessary to be able to participate in these forums, as well as the behavioral and socialization schemes that form part of daily social practices. The goal is to develop new methodologies for digital education that are more open and natural (Avila & Zacher-Pandya, 2012; Dominguez, 2006; Epstein, Nisbet & Gillespie, 2011; Hargittai, 2010).

Along with this first approach towards the active and applied component of web skills, the analytical concepts that underlie our study are the following:

- *The user is active on the web.* Web literacy is not only about “reading” the Web, but also “writing” on the Web. This includes reading, using the medium, and constructing online spaces where social practices take place. This does not mean that everyone must become a Webmaker, but it does mean that literate people should have the basic skills for connecting with others in order to interact both with and on the Web (Belshaw, 2013).

- *Skills and competencies.* A skill is a controlled activity that a person has learned to do. There are generic skills as well as specific ones. For example, a generic skill could be understanding how code is structured. A specific skill could be knowing how to use a diverse set of HTML elements. Skills have objective thresholds: they can be confirmed via evidence indicating that a predetermined skill level has been reached in a particular field. A competency, on the other hand, is a group of skills required for a predefined purpose. In spite of appearing objective, assessments of competencies are subjective by nature. These assessments require the prior definition of the criteria for considering a person “competent” in a particular subject and context (Belshaw, 2012).

- *Social web skills.* We were primarily interested in the access and use of the Web through browsers. There are “beginner” skills, such as identifying the web browser’s address bar, using functions like copy/paste, and directly including a site’s URL instead of searching for the page with a search engine. There are also “advanced” web skills, such as understanding code review workflow or server technologies. Both beginner skills as well as advanced ones lie outside the scope of this study.

Heuristic frameworks and human behaviors on the Web

We present a point of view that differs from classic conceptual approaches used to studying the world of web skills. It involves explicitly considering the heuristics and linking the practices of individuals using the web and the skills they develop from these practices.

The reasons for using a heuristic approach are twofold. First, the ability of subjects to act has expanded as a consequence of the massive use of digital technologies. This, in turn, has overrun the analytical frameworks that have traditionally explained the acquisition of technological skills as an isolated process. The frontiers between web literacy and social literacy are becoming more diffuse, as the digital realm massively permeates the rules of social conduct and shapes a model of network connectivity (Rainie & Wellman, 2012). It is becoming inappropriate to speak about digital competencies in education as something specific. The social practices of students who are connected daily through social network sites must also be considered (Ellison & Boyd, 2013).
Second, advances in the behavioral sciences have opened up new paths for researching how skills related to daily tasks are mastered. Notably, rationality theories have moved away from formal models towards action and an attempt to explain people’s decisions about the actions they will carry out. These theories explain that, in natural settings, there are many variables affecting the decision-making process. Under these circumstances, people apply mechanisms, termed heuristics, to adapt to environmental requirements and exploit them for their own benefit (De Neys, 2015; Evans, 2006; Gigerenzer & Gaissmaier, 2011; Goldstein, & Gigerenzer, 2002; March, 2002).

In this study, the characterization of heuristics is based on the concept of bounded rationality (Gigerenzer & Selten, 2001; Gilovich, Griffin, & Kahneman, 2002; Kahneman, 2003; Klaes & Sent, 2005) which lies within the field of cognitive science (Hutchins, 1995). According to Gerd, Gigerenzer & Reinhard Selten (2001), under the system of bounded rationality, actions are rational in proportion to the ability of the individual to exploit the structure of the action in which they find themselves, while at the same time taking into account the contextual and informational restrictions in which the action is carried out. Heuristics are “cognitive shortcuts” that allow individuals to evaluate a situation based on one or more basic rules. Using these, an individual can avoid carrying out costly, exhaustive evaluations of a varied and complex group of options. At the same time, the individual can pay attention to the circumstances that arise in that context. The advantage of bounded rationality and heuristics is that both theories elegantly explain how people can find suitable options in a way that is quick, successful, and a function of their environment. When a subject uses a heuristic he/she saves the effort of looking, identifying, and evaluating all the possible options in a context where these options are not readily apparent (Robles, 2007).

In education, heuristics have frequently been used as a methodological resource in disciplines that demand a high cognitive and meta-cognitive capacity (e.g., as applied to mathematics: Chavez, 2007; Hoon, Kee & Singh, 2013). Learning theories usually rely on analytical models and frameworks as a way of “operatizing” the variables related to a subjects’ behavior and explaining how these variables function in specific situations. For example, in the specific field of connected and online learning, heuristics and frameworks have been used on many occasions to analyze and interpret the processes and pathways governing the behavior of students in digitally mediated scenarios (Aparicio, Bacao & Oliveira, 2016; Conole, Galley, & Culver, 2011; Garrison & Anderson, 2003; Hirumi, 2002; Hwang et al., 2010; Wang, Chen & Anderson, 2014; Wang, Han & Yang, 2015).

Heuristics are not “a priori” constructions. Their identification emerges from practices that are often carried out daily in different situations. In the case of the Web, these practices occur in the context of situations such as identifying where to write a search term in a browser or learning how to edit information using an online word processor. These skills are considered prerequisites for developing competencies such as “Search Engine basics” and “Browser basics.” As a result of this decanting process, the heuristic that is linked to web literacy may consist of a series of competencies such as “Browser basics,” “Search engine basics,” and “Web mechanics.” By implication, the more highly-literate users will be those who automate their behavior on the Internet by relying on these competencies.

**Open and non-formal learning practices**

Open online education is a combination of the use of resources and tools under open licenses and the application of social media to promote learning. In a bottom-up model, like the one presented in this study, web skills are defined based on people’s practices in online spaces. Within these models, open learning contexts are considered a favorable forum in which to develop these skills.

In the social sphere, open learning has progressed due to the generalized adoption of the Internet by large, self-organized communities of civil society. These communities share resources and apply management methods based on connectivity. These communities serve different purposes (Carfagna, 2014). In the field of education, advances in open learning networks are facilitated by social media platforms, Massive Open Online Courses (MOOCs) and the Open Educational Resources movement. These advances place open education within a wider movement that drives the dissemination of knowledge in a way that is open, scalable, and capable of reaching large groups of society (Farrow et al., 2015; Weller, 2014; Wiley, 2014).

Open education projects allow any person to participate (within limits) regardless of their origin, location, or credentials. As in the case of traditional non-formal education, these projects challenge the notion of formal certification by “experts.” In these forums, new methods for evaluating learning outcomes in open online communities are utilized. Examples of these methods are the use of peer evaluation, the management of
reputation, complex meritocratic arrangements, and the measure of social capital within the community (Schmidt et al., 2009).

The open education approach challenges the way we think about education, educational institutions, and the certification of knowledge (Iiyoshi & Kumar, 2008). A “learning environment” no longer means a classroom or delimited digital zones, but rather encompasses many other wider spaces on the Internet. These spaces are distributed, connected without being limited by time and space (Gil-Jaurena & Domínguez, 2012).

The theory of Connected Learning specifically analyzes these situations and posits that, in connected environments, it is the students, who construct pathways to learning in a self-directed way. They do this by appropriating the resources that are most suited to their needs (Ito et al., 2013; Ito et al., 2015). Our research is consistent with this richer vision of learning. In this vision, it is possible for students to access the web and gain knowledge (Buckingham, 2007; Collins & Halverson, 2009; Williams, Karousou & Mackness, 2011). What matters is the ability of each individual to develop competencies through their own social behavior on the web. By applying these competencies in other situations, they achieve a level of web literacy (Dominguez & Trillo, 2014).

It is advantageous to use behaviors in open spaces as a model for understanding how Internet skills are acquired. According to Latchem (2014), there is still much to be learned about open education and how best to validate these skills using the existing evaluation and certification systems. The aim of this study is to understand this topography better by collecting information about those who develop competencies for web mastery through their practices in open spaces. Our objective is not to construct a model of competencies in order to create a standard for non-formal students. This approach has been criticized as lacking in empirical support. Categorization would also not be possible without prior research into how these skills are acquired by non-formal students. Categorization would require specifying the objectives of their behaviors and what processes and learning styles develop during skill acquisition (Coffield et al., 2004).

**Research context and methodology**

In order to validate the approach, a specific analysis was performed on data from a research study on the practices of university students in open learning spaces. Based on this research, a meta-analysis of the linkages and groupings between the student practices in these spaces was carried out in terms of the Web Literacy Map framework. We considered how they used the web and the skills they acquired from these practices.

The benchmark research consisted of two phases. The first phase was a case study of 40 university students enrolled in Communication Sciences (at the Complutense University of Madrid, Spain) and Social Education (at the National Distance Education University, UNED, Spain). The participants were offered an opportunity to take one of the open courses indexed in the Open Education Consortium website (www.oeconsortium.org/). This group was divided into five focus groups whose participants talked about their practices on the Web.

The second phase was an attempt to give the data greater representativity. The sample was widened to include 451 students with characteristics similar to members of the pilot group. The students were enrolled at the following universities: Granada University (Spain), Oviedo University (Spain), Vigo University (Spain), Complutense University of Madrid (Spain) and National Distance Education University (UNED, Spain). They completed a questionnaire-based survey that was organized in two dimensions. This was based on a contrast between the theoretical-conceptual aspects of the study and conclusions from the focus groups: (1) Generic Web Practices (9 items), (2) Learning Practices on the Web (10 items).

In turn, Mozilla’s Web Literacy Map was used as the heuristic around which the skills could be organized. The Web Literacy Map consisted of three dimensions, “Explore,” “Connect” and “Build.” However, the “Build” dimension and “Privacy” item of the “Connect” dimension were discarded because they only apply in the context of using the Web for learning purposes:

**Explore — Reading the Web**

- **Navigation.** Using software tools to browse the Web.
- **Web Mechanics.** Understanding the Web ecosystem and Internet stack.
- **Search.** Locating information, people and resources via the Web.
- **Credibility.** Critically evaluating information found on the Web.
- **Security.** Keeping systems, identities, and content safe.
Connect — Participating on the Web
- Sharing. Providing access to Web resources.
- Collaborating. Creating Web resources with others.
- Community participation. Getting involved in Web communities and understanding their practices.
- Open practices. Helping to keep the Web democratic and universally accessible.

Data analysis and interpretation

The meta-analysis began with a descriptive statistical analysis. This was followed by an exploratory factor analysis contrasting the relationships between web practices and the skills acquired. These skills were organized in a heuristic framework. The goal was to find parallelisms between the internal structure of the practices and their agreement with the heuristic of the Web Literacy Map. Therefore, we aimed to study the existing correlation between the structure of the factors derived from the practices and the structure of the Web Literacy Map.

Dimension 1 - Generic web practices

The structure of the items in this dimension is dichotomous. Its content matches the skills that fall into the “Explore” and “Connect” dimensions of the Web Literacy Map (see Table 1).

<table>
<thead>
<tr>
<th>Generic web practices</th>
<th>No (%)</th>
<th>Yes (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community participation</td>
<td>24.4</td>
<td>75.6</td>
</tr>
<tr>
<td>Navigation</td>
<td>25.3</td>
<td>74.7</td>
</tr>
<tr>
<td>Open practices</td>
<td>39.3</td>
<td>60.7</td>
</tr>
<tr>
<td>Sharing</td>
<td>42.8</td>
<td>57.2</td>
</tr>
<tr>
<td>Search</td>
<td>45.8</td>
<td>54.2</td>
</tr>
<tr>
<td>Web mechanics</td>
<td>59.7</td>
<td>40.3</td>
</tr>
<tr>
<td>Collaborating</td>
<td>60.6</td>
<td>39.4</td>
</tr>
<tr>
<td>Security</td>
<td>62.5</td>
<td>37.5</td>
</tr>
<tr>
<td>Credibility</td>
<td>74.2</td>
<td>25.8</td>
</tr>
</tbody>
</table>

Given this descriptive data, an exploratory factor analysis was applied. A KMO value of 0.78 was calculated. The Bartlett’s Test of Sphericity resulted in a score of 658. The significance level of 0.000 indicated the adequacy of the dimension reduction model. The variance explained by the four first factors was 68%. This resulted in the following groupings (see Table 2):

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>.101</td>
<td>.054</td>
<td>.156</td>
<td>.914</td>
</tr>
<tr>
<td>Open Practices</td>
<td>.057</td>
<td>.902</td>
<td>-.018</td>
<td>.044</td>
</tr>
<tr>
<td>Security</td>
<td>.713</td>
<td>.020</td>
<td>.010</td>
<td>.228</td>
</tr>
<tr>
<td>Sharing</td>
<td>.359</td>
<td>.176</td>
<td>.707</td>
<td>.091</td>
</tr>
<tr>
<td>Credibility</td>
<td>.835</td>
<td>.138</td>
<td>-.031</td>
<td>-.027</td>
</tr>
<tr>
<td>Collaborating</td>
<td>.526</td>
<td>.465</td>
<td>.308</td>
<td>-.161</td>
</tr>
<tr>
<td>Community Participation</td>
<td>-.269</td>
<td>-.170</td>
<td>.817</td>
<td>.083</td>
</tr>
<tr>
<td>Web Mechanics</td>
<td>.537</td>
<td>.430</td>
<td>.032</td>
<td>.260</td>
</tr>
<tr>
<td>Search</td>
<td>.412</td>
<td>.465</td>
<td>-.176</td>
<td>.385</td>
</tr>
</tbody>
</table>

Note. Rotation converged in 6 interactions. Extraction Method: Principal Component Analysis; Rotation Method: Varimax with Kaiser Normalization.

When assessing the variable factor saturations the factors listed here were revealed. These factors were consistent with the dimensions of the Web Literacy Map framework:
- The first factor was a general one related to practices of “Security,” “Credibility,” and “Web mechanics.” All of these were practices that formed part of the “Explore” dimension of the framework.
- The second factor was “Open Practices,” which fit into the “Connect” dimension of the framework.
• The third consisted of the practices of “Sharing” and “Community Participation,” in the “Connect” dimension of the framework.

• The fourth contained “Navigation,” of the “Explore” dimension of the framework.

**Dimension 2 - Learning practices on the Web**

The items structure in this dimension was designed in the form of a Likert-type scale with five levels (see Table 3).

<table>
<thead>
<tr>
<th>Table 3. Descriptive statistics, dimension: Learning practices on the Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic web practices</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Social media</td>
</tr>
<tr>
<td>Chat</td>
</tr>
<tr>
<td>Leisure</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td>Search of Non-academic Information</td>
</tr>
<tr>
<td>Exchange of Notes</td>
</tr>
<tr>
<td>Search of Academic Information</td>
</tr>
<tr>
<td>Readings</td>
</tr>
<tr>
<td>Performing Group Work (e.g., Skype)</td>
</tr>
<tr>
<td>Study tasks</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>N (%)</td>
</tr>
<tr>
<td>R (%)</td>
</tr>
<tr>
<td>S (%)</td>
</tr>
<tr>
<td>F (%)</td>
</tr>
<tr>
<td>A (%)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>14.8</td>
</tr>
<tr>
<td>5.3</td>
</tr>
<tr>
<td>11.9</td>
</tr>
<tr>
<td>24.0</td>
</tr>
<tr>
<td>44.0</td>
</tr>
<tr>
<td>19.4</td>
</tr>
<tr>
<td>8.1</td>
</tr>
<tr>
<td>10.3</td>
</tr>
<tr>
<td>22.1</td>
</tr>
<tr>
<td>40.2</td>
</tr>
<tr>
<td>19.0</td>
</tr>
<tr>
<td>9.7</td>
</tr>
<tr>
<td>16.4</td>
</tr>
<tr>
<td>24.1</td>
</tr>
<tr>
<td>30.8</td>
</tr>
<tr>
<td>20.8</td>
</tr>
<tr>
<td>9.3</td>
</tr>
<tr>
<td>15.8</td>
</tr>
<tr>
<td>17.2</td>
</tr>
<tr>
<td>37.0</td>
</tr>
<tr>
<td>20.9</td>
</tr>
<tr>
<td>13.9</td>
</tr>
<tr>
<td>24.1</td>
</tr>
<tr>
<td>21.8</td>
</tr>
<tr>
<td>19.3</td>
</tr>
<tr>
<td>53.4</td>
</tr>
<tr>
<td>14.5</td>
</tr>
<tr>
<td>15.2</td>
</tr>
<tr>
<td>9.7</td>
</tr>
<tr>
<td>7.1</td>
</tr>
<tr>
<td>39.5</td>
</tr>
<tr>
<td>18.3</td>
</tr>
<tr>
<td>18.4</td>
</tr>
<tr>
<td>14.6</td>
</tr>
<tr>
<td>9.3</td>
</tr>
<tr>
<td>50.4</td>
</tr>
<tr>
<td>16.4</td>
</tr>
<tr>
<td>17.1</td>
</tr>
<tr>
<td>9.0</td>
</tr>
<tr>
<td>7.2</td>
</tr>
<tr>
<td>64.9</td>
</tr>
<tr>
<td>15.8</td>
</tr>
<tr>
<td>8.6</td>
</tr>
<tr>
<td>4.1</td>
</tr>
<tr>
<td>6.6</td>
</tr>
<tr>
<td>57.2</td>
</tr>
<tr>
<td>23.9</td>
</tr>
<tr>
<td>9.8</td>
</tr>
<tr>
<td>4.9</td>
</tr>
<tr>
<td>4.3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Note.</strong> N = Never; R = Rarely; S = Sometimes; F = Frequently; A = Always.</td>
</tr>
</tbody>
</table>

Using this descriptive data, the exploratory factor analysis resulted in a KMO value of 0.82 and a value of 1069 on the Bartlett’s Test of Sphericity. The significance level was 0.000. These results suggested a dimension reduction model. This model consisted of three factors that accounted for 68% of the variance (see Table 4).

<table>
<thead>
<tr>
<th>Table 4. Factor Analysis, dimension: Learning practices on the Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotated component matrix*</td>
</tr>
<tr>
<td>Component</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>Study Tasks</td>
</tr>
<tr>
<td>Readings</td>
</tr>
<tr>
<td>Search of Non-academic Information</td>
</tr>
<tr>
<td>Chat</td>
</tr>
<tr>
<td>Social Media</td>
</tr>
<tr>
<td>Exchange of Notes</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td>Search of Academic Information</td>
</tr>
<tr>
<td>Performing Group Work (e.g., Skype)</td>
</tr>
<tr>
<td>Leisure</td>
</tr>
<tr>
<td>.789</td>
</tr>
<tr>
<td>.791</td>
</tr>
<tr>
<td>.098</td>
</tr>
<tr>
<td>.195</td>
</tr>
<tr>
<td>.081</td>
</tr>
<tr>
<td>.754</td>
</tr>
<tr>
<td>.254</td>
</tr>
<tr>
<td>.304</td>
</tr>
<tr>
<td>.757</td>
</tr>
<tr>
<td>.107</td>
</tr>
<tr>
<td>.122</td>
</tr>
<tr>
<td>.132</td>
</tr>
<tr>
<td>.449</td>
</tr>
<tr>
<td>.807</td>
</tr>
<tr>
<td>.873</td>
</tr>
<tr>
<td>.131</td>
</tr>
<tr>
<td>.208</td>
</tr>
<tr>
<td>.637</td>
</tr>
<tr>
<td>.131</td>
</tr>
<tr>
<td>.088</td>
</tr>
<tr>
<td>.661</td>
</tr>
<tr>
<td>.326</td>
</tr>
</tbody>
</table>

**Note.** Rotation converged in 5 interactions. Extraction Method: Principal Component Analysis; Rotation Method: Varimax with Kaiser Normalization.

Evaluating the factor saturations of the variables the following factors were found, which were consistent with the dimensions of the Web Literacy Map framework:

• The first factor was related to the use of the Web in academic activities, such as “Study tasks,” “Readings,” “Exchange of Notes,” and “Performing Group Work.” These were typical of the “Connect” dimension of the framework.

• The second factor was related to leisure and social interaction activities. These included “Chat,” “Social media,” and “Leisure.” These practices fit into the “Connect” dimension of the framework.

• The third includes the practices of “Search of Academic & Non-academic Information.” These practices fit into the “Explore” dimension of the framework.

**Discussion**

The current results should be interpreted keeping in mind that this was an analysis of students’ practices in open learning situations on the Internet. It is also important to consider that the Internet is a medium that provides...
access to new knowledge without restrictions from the physical world. In this forum, the student’s autonomy comes first. How well students shape their own educational environment is a subject for debate. This is especially true considering the emergence of social media and MOOCs. For example, Turkle (2011) has argued that young people who show over-confidence in social media are moving away from the meaning of social norms and the significance of conventional socialization. Other studies suggest that participation in digital media, especially multitasking, correlate with a reduction in sustained and reflective thought (Bauerlein, 2008; Baron, 2008; Carr, 2010; Greenfield, 2009; Pea et al., 2012). However, defenders of connected learning and new forms of endoculturation via digital media have argued in favor of types of learning that are highly-participative and that offer resources for digital self-literacy (Boyd, 2014; Jenkins, Ito, & Boyd, 2016). For example, Ito et al. (2008; 2013) analyzed the ways in which young people integrate diverse technologies into their lives for personal, social, and educational purposes. This group concluded that different groups of skills, literacies, and social relationships are needed in the connected world in order to prosper and be successful. These skills are acquired in open online environments.

Meanwhile, in accordance with bounded rationality theory, the intelligent use of heuristics can be utilized to adapt to the structure of the environment and exploit its resources in a way that favors certain behaviors. Application of this theory has given verifiable results in the fields of economics (Gilovich et al., 2002) and health (Thaler & Sunstein, 2008). Our work took these advances into consideration. We propose their application in the area of learning. Heuristics should be used as frameworks to guide methodologies in educational processes for web literacy. Still, what applications the heuristics model is appropriate is controversial subject. The definition of heuristics is not a deductive process, but rather an inductive one. Having to define the psychological mechanisms governing an individual’s rational actions complicates the task. This is the reason why bounded rationality is still not a complete and coherent corpus, but rather a “robust grouping of micro-theories” (Robles, 2007). The inclusion of experimental techniques in bounded rationality has led to advances in the field (Gigerenzer, 2015). Nevertheless, it is necessary to use empirical methods more often.

Data collection in non-formal environments is also discussed in the methodological section. There were unique challenges in the sampling and validation steps. This was due to the difficulties related to verification. In order to validate this type of methodology, Cohen (2007) conducted a multi-dimensional analysis of the research literature published in the areas of non-formal and formal education. He concluded that the usual qualitative and quantitative tools already available to researchers, such as interviews, observation, surveys, etc., are useful and appropriate for the study of non-formal education. He also concluded that the new tools and approaches were not needed to produce significant results. Although it can be scarcely found in the scientific literature, some evidence coming from analysis of the various research studies underscores the central role of non-formal and informal education. This evidence supports its relevance as a field of study in research of this kind (Latchem, 2012; Latchem, 2014).

**Conclusions**

The classical approach to digital literacy is the reference framework for web literacy. This approach assumes that digital skills are useful in order for people to be capable of selecting, analyzing, processing, organizing, and transforming information into knowledge based on context and personal and social needs. We believe that this approach is excessively instrumental. This is because it does not take into account the new competencies the web offers for people to be active in constructing new pathways for social participation and, especially, learning. In addition, incorporating a group of skills related to reading and writing on the web into the web literacy scheme is recommended. This group of skills should be included taking into account the Mozilla Foundation’s Web Literacy Map, used as the heuristic framework in this research.

The research data links students’ web practices skills developed via these practices. The organization of these practices is consistent with a heuristic framework that is based on the user’s behavior. The heuristic approach was applied in order to group the skills that were revealed while observing people’s web practices. Heuristics are constructed according to behavioral patterns. Heuristics are useful because grouping the practices seen in the observed behaviors facilitates the performance of complex tasks. This approach was used in this study to identify groups of web practices that have factors in common. In addition, these groups of practices are consistent with the skills defined under the framework of the Web Literacy Map. This framework is also a kind of heuristic model that is primarily defined on the basis of practices.

In conclusion, this practices-based-approach presents an alternative to classical web literacy instruction, which is primarily based on the conceptualization of a set of skills that need to be developed. The framework of this
research might also be useful for analyzing the field of web literacy and for composing educational proposals that take into account this practical dimension of social interactions.

References


**Effects of Attention Cueing on Learning Speech Organ Operation through Mobile Phones**

**Hui-Yu Yang**

Foreign Language Department, Fujian University of Technology, Fujian, China // yangh936@163.com

(Submitted March 14, 2016; Revised October 18, 2016; Accepted January 1, 2017)

**ABSTRACT**

The studies regarding using a cross sectional view of speech organs enriched with attention cueing and written text to probe learners’ learning efficiency and behavior through mobile phones is scant. The purpose of this study was to examine whether the presence of attention cueing can benefit learners with different amounts of prior knowledge in learning operational functions of the speech organs. Additionally, this study investigated the interactive effects of the experimental treatment and the learners’ prior knowledge on their test performance and cognitive load. A self-regulatory, mobile-phone-based text with accompanying pictures depicting places and manner of articulation comprised the instrumentation. The participants were comprised of 101 English as a foreign language (EFL) learners from four sections of a phonetics course. First, their level of prior knowledge was assessed. Next, they were randomly assigned to one of four modules—picture-only, picture-plus-signal, picture-plus-text, or picture-plus-text-plus-signal. Immediately after the treatment period, the participants were administered retention and transfer tests as well as cognitive load measurement. The results indicated that the enriched visualization somewhat reduced the participants’ cognitive load and enhanced their learning efficiency. However, presenting too much visual information impaired learning. The results of the study implied that enriched visualization compacted on the small screen might cause interference.

**Keywords**

Cognitive load, Mobile learning, Signaling principle, Working memory

**Introduction**

The prevalence of mobile phones enables users to obtain information anytime and anywhere due to their potential to make learning spontaneous, authentic, informal, situational, and personalized (Kukulska-Hulme, 2009). Today’s advanced digital functions in mobile phones along with wireless technology provide users with rich opportunities to access the real world, while allowing educators to extend learning beyond the traditional classroom (Thornton & Houser, 2005). Mobile phones have been employed as a language learning device to support the development of vocabulary (Chen, Hsieh, & Kinshuk, 2008), listening (Hsu, Hwang, Chang, & Yang, 2013), speaking (Ahn & Lee, 2010), reading (Hsu, Hwang, & Chang, 2013a) skills. However, due to mobile phones’ small screen sizes, incorporating mobile technology with alternative visualization options deserves further exploration (Chen et al., 2008; Kim & Kim, 2012).

According to the cognitive view of multimedia learning, the simultaneous presence of text and graphics will overburden a learner’s visual channel, whereby written text and images compete for the learners’ limited cognitive resources in processing the input and categorizing it in visual working memory. Simultaneously presenting text and images thus increases learners’ cognitive load and impairs learning. On the other hand, Baddeley’s (1986) working memory model claims that visual information is processed in the visuospatial sketchpad while verbal input is processed in the phonological loop. According to this view, presenting text and images simultaneously does not overload learners but rather helps them build up referential connections between text and images and achieve greater learning efficiency—the same concept that is behind dual coding theory (Paivio, 1986). Therefore, the explanation of Baddeley’s working memory model appears to be inconsistent with that of the cognitive view of multimedia learning.

Learners who visually search information using instructional visualizations may miss the focus. Providing visuospatial cueing on instructional visualization is expected to draw learners’ attention toward essential information, reduce visual search processes, foster selection of relevant information, and integrate information into a coherent mental representation (de Koning et al., 2009). However, whether the presence of enriched visualization (e.g., attention cueing and additional verbal information) can induce greater learning outcomes still remains controversial (de Koning et al., 2010b; de Koning et al., 2011a; Imhof et al., 2013).

EFL learners in China/Taiwan might be interfered by their mother tongue, use mother tongue to produce English sounds, and result in incorrect pronunciation. Besides, EFL learners cannot observe the operational mechanism of speech organs whenever they produce sounds, to figure out the operation of speech organs appears to be too
abstract for them. Therefore, presenting EFL learners with a cross sectional view of speech organs by showing
the operation of tongue, velum and vocal folds will help them understand the mechanism of sound production.
Besides, the addition of motion-indicating arrows to indicate the movement of speech organs will help learners
understand an abstract concept and build up a mental model (Hegarty, Kriz, & Cate, 2003).

While incorporating learning theories and instructional visualization with mobile phone technology in examining
learners’ learning outcomes, individual learner characteristics should also be considered (Chen et al., 2008). This
study explores the effects of different forms of visualization and learners’ expertise level on construction of
mental models (Hegarty et al., 2003) and cognitive load in learning the operation of human speech organs
through mobile phones.

Literature review

Mobile assisted language learning

Previous studies have investigated the use of handheld devices in supporting foreign language learners’
development of vocabulary (e.g., Chen et al., 2008; Kim & Kim, 2012; Sandberg et al., 2014; Thornton &
Houser, 2005), listening (e.g., Hsu et al., 2013b), speaking (e.g., Ahn & Lee, 2016; Liu & Chu, 2010), and
reading (e.g., Hsu et al., 2013a).

Chen et al. (2008) presented four different annotations to probe vocabulary acquisition among EFL learners with
different visuospatial aptitudes, and found that they learned better when the annotation type suited their
visuospatial aptitudes. Sandberg et al. (2014) designed a gaming context and intelligent adaptation mode for ESL
children to learn animal-related vocabulary, and noted the positive impacts of using adventure games in
children’s vocabulary acquisition. Thornton and Houser (2005) conducted a study in which they sent English
vocabulary and idioms via email to Japanese university students’ mobile phones, and observed that doing so
prompted learners to review the contents more and also resulted in better learning outcomes compared with
students who learn the vocabulary and idioms through web- and paper-based materials.

Hsu et al. (2013b) presented EFL learners with three types of movie captions through handheld devices and
supported the potential of mobile devices for vocabulary acquisition and listening skill development. Liu and
Chu (2010) constructed a ubiquitous game-based learning environment to help EFL learners develop listening
and speaking skills through Personal Digital Assistant (PDA) device, in which gaming users showed more
favorable learning outcomes and enhanced motivation compared with non-gaming users. Ahn and Lee (2016)
employed speech recognition technology to assist EFL learners in practicing speaking. The learners gave positive
comments, suggesting that mobile phones have potential to support the development of speaking skills in EFL
contexts.

Hsu et al. (2013a) designed a recommendation-based system that involved recommend reading materials tailored
for individual EFL learners based on their English proficiency and reading preferences. The system required
learners to take notes in either individual- or shared-annotation mode, which they felt comfortable and satisfied
with. This system also helped improve learners’ reading comprehension.

Empirical studies concerning mobile learning have highlighted the potential of mobile devices on language skill
development. However, learning the operation of speech organs to understand the pronunciation of IPA symbols
through mobile phones was not probed in the past. Because small screens might be challenging for learners to
perceive visualization (Stockwell, 2008; Stockwell, 2012; Thornton & Houser, 2005) and result in perceptual
errors (Kim & Kim, 2012); presenting attention cueing in mobile-phone-based visualization might reduce
learners’ cognitive load and enhance learning efficiency (Liu, Lin, & Paas, 2013).

Signaling principle

In terms of perceptual processing, the addition to the visualizations that contrasts explicitly with corresponding
features of its neighbors can preferentially attract learners’ attention and ameliorate the processing cost
associated with complex visualizations (Lowe & Boucheix, 2011).

Crooks et al. (2012) and Ari et al. (2014) examined the effects of cueing and modality on a computer-based
diagram depicting places of articulation in human speech. They found that the presence of cueing yielded no
significant differences between learners in terms of their test results and cognitive load ratings. A study conducted by Lin and Atkinson (2011) also indicated that learners’ test performance and cognitive load were similar regardless of whether or not visual cues were provided. However, visual arrow cueing reduced learners’ study time on the instructional visualizations. Boucheix and his colleagues (2013) conducted studies concerning learning the operation of a piano mechanism and found the localized coordinate and progressive path cueing yielded a better idea about the operation of a piano mechanism compared with entity cueing and non-cueing. The spreading-color cueing presented the learners with specific and continuous spatio-temporal direction of attention by establishing pathway of causal chains; however, arrow cueing failed to establish continuous pathway of spatial and temporal linkage for each entity and resulted in unfavorable mental model development (e.g., Boucheix & Lowe, 2010). While another similar study observed no promising effects of visuospatial cueing (e.g., Lowe & Boucheix, 2011).

Imhof et al. (2013) investigated the effects of multiple vs. single visualizations with the presence/absence of arrow cueing on learning fish locomotion patterns. The experimental results indicated that enriched visualizations with attention cueing caused interference, wherein simultaneous highlighting of multiple elements without specificity caused confusion (Moreno, 2007). Kriz and Hegarty (2007) speculated that attention cueing aims to capture learners’ attention and directs it toward specific or essential elements in visualizations but without conveying any new information or guaranteeing that conceptual understanding and construction of mental models have taken place.

de Koning and his colleagues (2011b) conducted several studies in which they decreased the luminance of uncued subsystems to show a visual contrast between them and cued subsystems in an animated cardiovascular system. The learners exhibited similar test performances regardless of cueing conditions and display speeds. Besides, studies addressing self- or instructional explanations accompanied with attention cueing (de Koning et al., 2010b; de Koning et al., 2011b) had yielded inconsistent results. A study by Kalyuga, Chandler, and Sweller (1999) demonstrated that the presence of enriched color cueing captured learners’ attention and yielded better efficiency when they had to split their attention to process text and diagrams.

In sum, studies investigating the promising effects of attention cueing on multimedia learning efficiency have yielded mixed results. The use of attention cueing has been shown to facilitate learning efficiency and reduce learners’ cognitive load (e.g., Boucheix & Lowe, 2010; Boucheix, et al., 2013; de Koning et al., 2010b; de Koning et al., 2011b; Imhof et al., 2013; Kalyuga et al., 1999). Other studies have shown that attention cueing may facilitate learning efficiency but without reducing cognitive load (de Koning et al., 2010b). Still other studies have found attention cueing to be ineffective and even causes interference (e.g., Crooks et al., 2012; de Koning et al., 2010a; de Koning et al., 2011a; Imhof et al., 2013; Kriz & Hegarty, 2007; Lin & Atkinson, 2011; Lowe & Boucheix, 2011; Moreno, 2007).

**Statement of the problem**

Some empirical researches concerning mobile learning did not incorporate a control group (Sandberg et al., 2014). Moreover, whether providing attention cueing can encourage learners to extract and process visual information and thereby build up mental models have shown inconsistent results. Research suggests that visual cueing will be more effective in static visualizations than in animation (Boucheix et al., 2013; de Koning et al., 2009). The quality of mental model construction through static visualizations can be improved by enriching the static visualizations with motion-indicating arrows (Heiser & Tversky, 2002) or written information (Hegarty & Just, 1993; Hegarty et al., 2003). Verbal text accompanying diagrams may vividly illustrate abstract concepts which may especially benefit novices (Kriz & Hegarty, 2007). The present research will investigate whether integrating both written information and attention cueing can augment learning results more than using only a single approach (e.g., Crooks et al., 2012). Besides, prior knowledge of how speech organs function through the use of mobile-phone-based visualization as a modulating factor was ignored in the past (e.g., Ari et al., 2014; Crooks et al., 2012). Learners’ study time and number of clicks have also not been probed in the past (Crooks et al., 2012). To address the unresolved issues, the research questions in the present study are as follows:

- Do learners perform better when given attention cueing compared to not having attention cueing when learning about the operation of the speech organs?
- Do learners who are given attention cueing experience a lower cognitive load when learning about the operation of the speech organs compared to learners who are not given attention cueing?
- Do learners study for different lengths of time and have a differing number of clicks under different presentation modes?
Methodology

Participants

A total of 101 EFL learners (male = 23, female = 78) with an age of 19 ($M = 19.29$, $SD = 0.91$) volunteered to participate in the experiment. They were recruited from four classes of introduction to phonetics course at a technology university in southeastern China. The participants had not had previous experience with the instrument used in the present study. According to the prior knowledge test score, the top 30% were classified into high-knowledge and the bottom 30% were low-knowledge learners. Each group was comprised of both high- and low-knowledge learners. A one-way ANOVA revealed no significant difference among the four groups with respect to their prior knowledge, $F(3,97) = 0.648$, $p = .586$.

Research design

The present study examined four conditions with the purpose of determining the potential benefit of enriched information contained in static visualizations by investigating whether motion-indicating arrows and written text in static visualizations were more beneficial than the same visualizations without arrows or written text. Presentation mode and prior knowledge were the independent variables. Retention and transfer test results, cognitive load ratings, study time, and number of clicks were the dependent variables (Figure 1).

Independent variables

Experimental treatment

The instrumentation (Figure 2) was developed by the researcher and administered to the students in the formal experiment. The picture was at the bottom, while each IPA (International Phonetic Alphabet) symbol was at the top of the picture. When a student clicked on an IPA symbol, a corresponding picture popped up, depicting the place and manner of articulation. A total of twenty-nine IPA symbols (i.e., $a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, \theta, \phi, \alpha, \beta, \gamma, \delta$) used to represent how the speech organs operate. A recorder was installed in the system to record how many times the students clicked on each IPA symbol and how much time they spent learning how the speech organs operate.

In the experiment, four presentation modes were designed respectively: picture-only, picture-plus-signal, picture-plus-text, and picture-plus-text-plus-signal. For the picture-only group (PG), the learners were given the IPA symbols and corresponding static visualizations only. When the students clicked on each IPA symbol, one static picture appeared on the screen showing static status of the lips shape, tongue, velum, and vocal folds when that
IPA symbol was vocalized (Figure 2a). For the picture-plus-signal group (PSG), the learners received the IPA symbols and corresponding static visualizations enriched with attention cueing. When the learners clicked each IPA symbol, one static picture embedded with attention cueing popped up showing the static positions of the lips shape, tongue, velum, and vocal folds. The picture was enriched with attention cueing when the speech organs had one of following features: (1) different shapes of lips when producing bilabials, labiodental-fricatives, or stops; (2) nasalization; (3) voiced sounds; and (4) the tongue in dental, alveolar, post-alveolar, palatal, velar, or uvular positions. For example, in the picture corresponding to [n], one motion-indicating arrow alerted the learners of the movement of the velum; and a wavy line demonstrated how the vocal folds vibrated (Figure 2b). On the other hand, the picture was not enriched with attention cues when the speech organs had one of following features: (1) spread lips; (2) oral sound; (3) voiceless sounds; and (4) the tongue in a resting status. For the picture-plus-text group (PTG), the learners were given the IPA symbols, corresponding static visualization and the verbal text. When the students clicked on each IPA symbol, one static picture appeared depicting the static positions of the speech organs enriched with verbal texts annotating places of articulation, lips shape, and status of velum and vocal folds (Figure 2c). In the picture-plus-text-plus-signal group (PTSG), the learners were given the same visualizations as those in the PTG but enriched with the same attention cueing as those in the PSG (Figure 2d).

Prior knowledge test

This was comprised of a table of 24 IPA symbols (i.e., p, b, t, d, k, g, f, v, s, z, l, s, f, s, z, n, w, l, r, h, th, j, b, and ð). The students chose the correct symbol from a drop-down menu (i.e., [p] should be selected from the menu under the column “voiceless, bilabial, stop”). Each correct answer gained one point (Figure 3).
Dependent variables

Retention test

The purpose of the test was to assess the learners’ understanding and retention of the instructional visualizations. The test consisted of 20 multiple-choice questions regarding the phonetic symbols (i.e., b, t, m, d, n, x, s, r, h, v, θ, ϕ, s, z, l, ð, θ, θ, and ѱ). The learners made inferences from the diagrams and verbal texts, and chose an IPA symbol (Figure 4a). Each correct answer gained one point.

Transfer test

The students’ application ability was assessed by 20 multiple-choice questions containing: (1) IPA symbols that are used in some languages (i.e., ś, ř, ż, r̃, ʃ, t̃, l̃, ñ, ñ, h̃, and ř̃); and (2) IPA symbols that are not regularly used in any languages (i.e., s, z, θ, ḟ, j̃, ṽ, and ř̃). These IPA symbols never appeared in the instrumentation, and the students had never encountered them in their daily life. The questions were to assess how much learners
could apply what they had learned from the instructional visualization, make inferences from the diagrams and verbal texts, and think about a possible IPA symbol (Figure 4b). Dr. Daniel Currie Hall was consulted about the usage of these IPA symbols in the world’s languages. Each correct answer gained one point.

**Cognitive load**

A subjective cognitive load measurement with a scale ranging from 1 to 9 (Kalyuga et al., 1999; Paas, 1992) measured the learners’ cognitive load, with 1 representing the lowest level and 9 representing the highest level. Item 1 dealt with intrinsic load; item 2 dealt with extraneous load; item 3 probed germane load; item 4 examined perceived difficulties in answering the retention test; and item 5 queried perceived difficulties in answering the transfer test.

**Study time and number of clicks**

A recorder installed in the system recorded the total amount of seconds the learners spent learning the operation of the speech organs, and the total number of clicks they clicked on each symbol.

**Data collection instrumentation**

A pilot study involving 63 undergraduates in the College of Humanities was conducted before the experiment. Point-biserial correlation was conducted to eliminate the less reliable test items (Wu & Tu, 2006). The reliability of each measurement is detailed below:

**Prior knowledge test, retention test, and transfer test**

After item analysis, all of the items on the prior knowledge test (Cronbach’s alpha = 0.909) were retained. Three questions on the retention test (Cronbach’s alpha = 0.764) and five questions on the transfer test (Cronbach’s alpha = 0.740) were eliminated. All thirty-two test items in the retention and transfer tests were preserved since they demonstrated good internal consistency (Cronbach’s alpha = 0.811).

**Cognitive load measurement**

A Pearson correlation analysis revealed a significant result on the critical ratio and item-total correlation. The Cronbach’s alpha was 0.859. Bartlett’s test of sphericity was significant and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) was 0.832. The eigenvalue was greater than 1 and the total explained variance was 86.962%.

**Experimental procedures**

The students were seated in a language laboratory and were required to use the 5-inch 4G mobile phones provided. Each mobile phone had a resolution of 720 x 1280. All instructional visualizations were conveyed through the mobile phones. The presentation mode comprised a self-regulatory model to allow the students to spend as much time as they wanted to learn the lessons and answer all the questions. If any of the students missed answering any question before moving on to next step, they would be reminded by the system to complete the missing information. The experimental procedures consisted of five stages:

First, the students were required to enter their name, age, gender, and student ID to log into the designated web page. Then they took a prior knowledge test concerning English phonetics.

Secondly, they moved on to read the explanations of the interface and experimental procedures on mobile phones silently by themselves before the start of the formal experiment.

Thirdly, in the formal experiment, when the students were learning about the operation of the speech organs, they were allowed to click each IPA symbol as many times as they wanted to learn the positions of lips shape, tongue,
velum, and vocal folds. Afterwards, they answered three cognitive load questions concerning the difficulty level of the instrumentation.

Fourthly, the students were given 2 minutes to read a brief introduction about interface features of the retention and transfer tests. Some terminology and particular IPA symbols were explained to reduce the learners’ anxiety regarding the tests. After 2 minutes, the system automatically initiated the retention test. The students took the retention test and submitted their answers. Then, they answered one cognitive load question concerning the retention test.

Finally, they took the transfer test and submitted their answers after completing 20 multiple-choice. Then, they answered one cognitive load question regarding the transfer test.

After answering all of the questions and submitting all answers into the system, the students logged out. The server logs recorded the students’ learning activities, study time, and test results. The overall experimental procedure took approximately 1.5 hours. The data from the four conditions were collected during separate class periods.

**Results**

**Research question one: Do learners perform better when given attention cueing compared to not having attention cueing when learning about the operation of the speech organs?**

A Pearson correlation revealed a statistically significant negative correlation between the learners’ test scores and overall cognitive loads, $r = -0.204, p = 0.041$. As the learners’ cognitive load decreased, their test performance increased, and vice versa.

A two-way ANOVA was conducted to examine the interactive effects between the experimental conditions and prior knowledge on the test results. In Table 1, the ANOVA source of variation results indicated no significant interaction effects on the retention test, $F(3,93) = 0.837, p = 0.477$, partial $\eta^2 = 0.026$. However, the main effect of the experimental condition was statistically significant, $F(3,93) = 5.146, p = 0.002$, partial $\eta^2 = 0.142$. The PG ($M = 11.50, SD = 3.934$) significantly outperformed the PTSG ($M = 8.48, SD = 2.517$), $p = 0.09$. The PSG ($M = 11.70, SD = 2.914$) significantly outperformed the PTSG ($M = 8.48, SD = 2.517$), $p = 0.005$. The main effect of prior knowledge was statistically significant, $F(1,93) = 4.556, p = 0.035$, partial $\eta^2 = 0.047$. The learners with high prior-knowledge ($M = 11.23, SD = 3.598$) significantly outperformed those with low prior-knowledge ($M = 9.83, SD = 3.418$) (see Table 1).

There were no interaction effects in the transfer test, $F(3,93) = 0.774, p = 0.511$, partial $\eta^2 = 0.024$. However, the main effect of the experimental condition was statistically significant, $F(3,93) = 2.808, p = 0.044$, partial $\eta^2 = 0.083$. The PSG ($M = 8.13, SD = 2.242$) significantly outperformed the PG ($M = 5.67, SD = 2.839$), $p = 0.016$. The main effect of prior knowledge was not statistically significant, $F(1,93) = 3.907, p = 0.051$, partial $\eta^2 = 0.040$.

<table>
<thead>
<tr>
<th>Group</th>
<th>Retention test</th>
<th>Transfer test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>PG (N = 24)</td>
<td>11.50</td>
<td>3.934</td>
<td>5.67</td>
</tr>
<tr>
<td>PSG (N = 23)</td>
<td>11.70</td>
<td>2.914</td>
<td>8.13</td>
</tr>
<tr>
<td>PTG (N = 27)</td>
<td>10.59</td>
<td>3.876</td>
<td>7.48</td>
</tr>
<tr>
<td>PTSG (N = 27)</td>
<td>8.48</td>
<td>2.517</td>
<td>6.78</td>
</tr>
<tr>
<td>High (N = 48)</td>
<td>11.23</td>
<td>3.598</td>
<td>7.65</td>
</tr>
<tr>
<td>Low (N = 53)</td>
<td>9.83</td>
<td>3.418</td>
<td>6.43</td>
</tr>
</tbody>
</table>

There were no interaction effects in total score $F(3,93) = 0.841, p = 0.475$, partial $\eta^2 = 0.026$. However, the main effect of the experimental treatment was statistically significant, $F(3,93) = 3.159, p = 0.028$, partial $\eta^2 = 0.092$. The PSG ($M = 19.83, SD = 4.428$) significantly outperformed the PTSG ($M = 15.26, SD = 4.528$), $p = 0.012$. The main effect of prior knowledge was also statistically significant, $F(1,93) = 6.013, p = 0.016$, partial $\eta^2 = 0.061$. The high prior-knowledge learners ($M = 18.88, SD = 5.541$) significantly outperformed the low prior-knowledge learners ($M = 16.26, SD = 5.016$) (see Table 1).
Research question two: Do learners who are given attention cueing experience a lower cognitive load when learning about the operation of the speech organs compared to learners who are not given attention cueing?

A two-way ANOVA was conducted to examine the interactive effects between the experimental conditions and prior knowledge on cognitive load. In Table 4, the ANOVA source of variation results indicated no significant interaction effects on intrinsic load, $F(3,93) = 1.167$, $p = .327$, partial $\eta^2 = .036$. The main effect of group treatment was not statistically significant, $F(3,93) = 2.216$, $p = .091$, partial $\eta^2 = 0.067$. The main effect of prior knowledge was statistically significant, $F(1,93) = 5.221$, $p = .026$, partial $\eta^2 = 0.053$. The low-knowledge learners ($M = 5.79, SD = 1.691$) had higher intrinsic load than the high-knowledge learners ($M = 5.02, SD = 1.564$) (see Table 3).

### Table 3. Cognitive load ratings under different conditions

<table>
<thead>
<tr>
<th>Group</th>
<th>Intrinsic</th>
<th>Extraneous</th>
<th>Germane</th>
<th>Retention</th>
<th>Transfer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>PG ($N=24$)</td>
<td>5.79</td>
<td>1.503</td>
<td>5.67</td>
<td>1.971</td>
<td>6.04</td>
<td>1.899</td>
</tr>
<tr>
<td>PSG ($N=23$)</td>
<td>5.26</td>
<td>2.072</td>
<td>4.96</td>
<td>1.581</td>
<td>4.96</td>
<td>1.551</td>
</tr>
<tr>
<td>PTG ($N=27$)</td>
<td>4.81</td>
<td>1.178</td>
<td>4.85</td>
<td>1.350</td>
<td>5.41</td>
<td>1.217</td>
</tr>
<tr>
<td>PTSG ($N=27$)</td>
<td>5.85</td>
<td>1.725</td>
<td>5.81</td>
<td>1.733</td>
<td>6.22</td>
<td>1.577</td>
</tr>
<tr>
<td>High ($N=48$)</td>
<td>5.02</td>
<td>1.564</td>
<td>5.02</td>
<td>1.564</td>
<td>5.33</td>
<td>1.562</td>
</tr>
<tr>
<td>Low ($N=53$)</td>
<td>5.79</td>
<td>1.691</td>
<td>5.60</td>
<td>1.780</td>
<td>5.98</td>
<td>1.635</td>
</tr>
</tbody>
</table>

There were no interaction effects in regard to extraneous load, $F(3,93) = 0.595$, $p = .620$, partial $\eta^2 = 0.019$. Neither the main effect of group treatment, $F(3,93) = 2.051$, $p = .112$, partial $\eta^2 = 0.062$, nor the main effect of prior knowledge was significant, $F(1,93) = 2.509$, $p = .117$, partial $\eta^2 = 0.026$.

There were not interaction effects in regard to germane load, $F(3,93) = 0.499$, $p = .684$, partial $\eta^2 = 0.016$. However, the main effect of group treatment was statistically significant, $F(3,93) = 3.114$, $p = .030$, partial $\eta^2 = 0.091$. The PTSG ($M = 6.22, SD = 1.577$) had a significantly higher germane load than did the PSG ($M = 4.96, SD = 1.551$), $p = 0.027$ (see Table 3). The main effect of prior knowledge was not statistically significant, $F(1,93) = 3.609$, $p = .060$, partial $\eta^2 = 0.037$.

There were no interaction effects in regard to retention load, $F(3,93) = 0.350$, $p = .790$, partial $\eta^2 = 0.011$. The main effect of group treatment was statistically significant, $F(3,93) = 9.451$, $p = .000$, partial $\eta^2 = 0.234$. The PG ($M = 7.83, SD = 1.465$) had a higher retention load than did those in the PSG ($M = 5.83, SD = 1.922$), $p = .000$ as well as those in the PTG ($M = 5.67, SD = 0.784$), $p = .000$ and in the PTSG ($M = 6.37, SD = 1.735$), $p = .006$ (see Table 3). The main effect of prior knowledge was not statistically significant, $F(1,93) = 0.216$, $p = .643$, partial $\eta^2 = 0.002$.

There were no interaction effects in regard to transfer load, $F(3,93) = 1.360$, $p = .260$, partial $\eta^2 = 0.042$. However, the main effect of group treatment was statistically significant, $F(3,93) = 6.095$, $p = .001$, partial $\eta^2 = 0.164$. The PG ($M = 8.29, SD = 1.301$) had a higher transfer load than the PSG ($M = 6.39, SD = 2.231$), $p = .000$.
as well as the PTG (M = 6.93, SD = 1.299), p = .014 (see Table 3). The main effect of prior knowledge was not statistically significant, F(1,93) = 0.122, p = .728, partial η² = 0.001.

### Table 4. Results of two-way ANOVAs on cognitive load

<table>
<thead>
<tr>
<th>Source</th>
<th>Load</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>η²p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Intrinsic</td>
<td>17.044</td>
<td>3</td>
<td>5.681</td>
<td>2.216</td>
<td>.091</td>
<td>.067</td>
</tr>
<tr>
<td></td>
<td>Extrinsic</td>
<td>17.080</td>
<td>3</td>
<td>5.693</td>
<td>2.051</td>
<td>.112</td>
<td>.062</td>
</tr>
<tr>
<td></td>
<td>Germane</td>
<td>22.834</td>
<td>3</td>
<td>7.611</td>
<td>3.114</td>
<td>.030</td>
<td>.091</td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>67.698</td>
<td>3</td>
<td>22.566</td>
<td>9.451</td>
<td>.000</td>
<td>.234</td>
</tr>
<tr>
<td></td>
<td>Transfer</td>
<td>45.479</td>
<td>3</td>
<td>15.160</td>
<td>6.095</td>
<td>.001</td>
<td>.164</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>626.667</td>
<td>3</td>
<td>208.889</td>
<td>5.520</td>
<td>.002</td>
<td>.151</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>Intrinsic</td>
<td>13.383</td>
<td>1</td>
<td>13.383</td>
<td>5.221</td>
<td>.025</td>
<td>.053</td>
</tr>
<tr>
<td></td>
<td>Extrinsic</td>
<td>6.966</td>
<td>1</td>
<td>6.966</td>
<td>2.509</td>
<td>.117</td>
<td>.026</td>
</tr>
<tr>
<td></td>
<td>Germane</td>
<td>8.847</td>
<td>1</td>
<td>8.847</td>
<td>3.620</td>
<td>.060</td>
<td>.037</td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>1.017</td>
<td>1</td>
<td>1.017</td>
<td>.216</td>
<td>.643</td>
<td>.002</td>
</tr>
<tr>
<td></td>
<td>Transfer</td>
<td>.303</td>
<td>1</td>
<td>.303</td>
<td>.122</td>
<td>.728</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>111.117</td>
<td>1</td>
<td>111.117</td>
<td>2.936</td>
<td>.090</td>
<td>.031</td>
</tr>
<tr>
<td>Group * knowledge</td>
<td>Intrinsic</td>
<td>8.971</td>
<td>3</td>
<td>2.990</td>
<td>1.167</td>
<td>.327</td>
<td>.036</td>
</tr>
<tr>
<td></td>
<td>Extrinsic</td>
<td>4.956</td>
<td>3</td>
<td>1.652</td>
<td>0.595</td>
<td>.620</td>
<td>.019</td>
</tr>
<tr>
<td></td>
<td>Germane</td>
<td>3.661</td>
<td>3</td>
<td>1.220</td>
<td>0.499</td>
<td>.684</td>
<td>.016</td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>2.504</td>
<td>3</td>
<td>1.835</td>
<td>0.350</td>
<td>.790</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>Transfer</td>
<td>10.151</td>
<td>3</td>
<td>3.384</td>
<td>1.360</td>
<td>.260</td>
<td>.042</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>105.360</td>
<td>3</td>
<td>35.120</td>
<td>.928</td>
<td>.430</td>
<td>.029</td>
</tr>
<tr>
<td>Error</td>
<td>Intrinsic</td>
<td>93</td>
<td></td>
<td>2.563</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extrinsic</td>
<td>93</td>
<td></td>
<td>2.776</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Germane</td>
<td>93</td>
<td></td>
<td>2.444</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>93</td>
<td></td>
<td>2.388</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transfer</td>
<td>93</td>
<td></td>
<td>2.487</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>93</td>
<td></td>
<td>37.844</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

There were no interaction effects in regard to overall cognitive load, F(3,93) = 0.928, p = .430, partial η² = 0.029. The main effect of group treatment was statistically significant, F(3,93) = 5.520, p = .002, partial η² = 0.151. The PG (M = 33.63, SD = 6.371) had a higher overall cognitive load than the PSG (M = 27.39, SD = 7.421), p = .004 as well as the PTG (M = 27.67, SD = 4.867), p = .005 (see Table 3). There was a significant trend in prior knowledge, F(1,93) = 2.936, p = .090, partial η² = 0.031.

**Research question three: Do learners study for different lengths of time and have a differing number of clicks under different presentation modes?**

A two-way ANOVA was conducted to examine the interactive effects between the experimental conditions and prior knowledge on study patterns. In Table 6, the ANOVA source of variation results indicated no significant interaction effects on study time F(3,93) = 1.814, p = .150, partial η² = 0.055. The main effect of group treatment was statistically significant, F(3,93) = 4.934, p = .003, partial η² = 0.137. Learners in the PTG (M = 408.37, SD = 411.297) spent more time than did those in the PSG (M = 176.13, SD = 134.290), p = .013 as well as those in the PTSG (M = 168.30, SD = 143.531), p = .006 (see Table 5). The main effect of prior knowledge was not statistically significant, F(1,93) = 0.043, p = .836, partial η² = 0.000.

### Table 5. Results of study pattern

<table>
<thead>
<tr>
<th>Group</th>
<th>Study time</th>
<th>Number of clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>PG (N = 24)</td>
<td>255.75</td>
<td>256.848</td>
</tr>
<tr>
<td>PSG (N = 23)</td>
<td>176.13</td>
<td>134.290</td>
</tr>
<tr>
<td>PTG (N = 27)</td>
<td>408.37</td>
<td>411.297</td>
</tr>
<tr>
<td>PTSG (N = 27)</td>
<td>168.30</td>
<td>143.531</td>
</tr>
<tr>
<td>High-knowledge</td>
<td>259.69</td>
<td>235.180</td>
</tr>
<tr>
<td>Low-knowledge</td>
<td>250.83</td>
<td>317.067</td>
</tr>
</tbody>
</table>

121
There were no interaction effects in regard to number of clicks, $F(3,93) = 0.868$, $p = .461$, partial $\eta^2 = 0.027$. The main effect of group treatment was statistically significant, $F(3,93) = 5.510$, $p = .002$, partial $\eta^2 = 0.151$. The PG ($M = 81.67$, $SD = 69.717$) significantly had more number of clicks than the PSG ($M = 22.78$, $SD = 26.831$), $p = .002$. The PTG ($M = 71.11$, $SD = 62.635$) significantly had more number of clicks than the PSG ($M = 22.78$, $SD = 26.831$), $p = .014$. The PTSG ($M = 70.22$, $SD = 48.875$) significantly had more number of clicks than the PSG ($M = 22.78$, $SD = 26.831$), $p = .017$ (see Table 5). The main effect of prior knowledge was not statistically significant, $F(1,93) = 0.051$, $p = .822$, partial $\eta^2 = 0.001$.

### Table 6. Results of two-way ANOVA on study pattern

<table>
<thead>
<tr>
<th>Source</th>
<th>Test</th>
<th>Type III Sum of Squares</th>
<th>$df$</th>
<th>$MS$</th>
<th>$F$</th>
<th>Sig.</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Study time</td>
<td>1029604.844</td>
<td>3</td>
<td>343201.615</td>
<td>4.934</td>
<td>.003</td>
<td>.137</td>
</tr>
<tr>
<td></td>
<td>Clicks</td>
<td>50422.072</td>
<td>3</td>
<td>168073.357</td>
<td>5.510</td>
<td>.002</td>
<td>.151</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>Study time</td>
<td>3014.436</td>
<td>1</td>
<td>3014.436</td>
<td>.043</td>
<td>.836</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Clicks</td>
<td>155.996</td>
<td>1</td>
<td>155.996</td>
<td>.051</td>
<td>.822</td>
<td>.001</td>
</tr>
<tr>
<td>Group * knowledge</td>
<td>Study time</td>
<td>378566.907</td>
<td>3</td>
<td>126188.969</td>
<td>1.814</td>
<td>.150</td>
<td>.055</td>
</tr>
<tr>
<td></td>
<td>Clicks</td>
<td>7939.852</td>
<td>3</td>
<td>2646.617</td>
<td>.868</td>
<td>.461</td>
<td>.027</td>
</tr>
<tr>
<td>Error</td>
<td>Study time</td>
<td>69562.421</td>
<td>93</td>
<td>69562.421</td>
<td>.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clicks</td>
<td>3050.597</td>
<td>93</td>
<td>3050.597</td>
<td>.93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

### Discussion

#### Research question 1

The presence of either cueing or verbal text was beneficial to help the learners develop greater conceptual understanding, construct mental models, and result in deep learning (Hegarty & Just, 1993; Heiser & Tversky, 2002). The results in the experiment (see Table 7) were partially consistent with the results of previous studies (e.g., Boucheix & Lowe, 2010; Boucheix, et al., 2013; de Koning et al., 2010b; de Koning et al., 2011b; Imhof et al., 2013; Kalyuga et al., 1999) in which attention cueing helped learners organize and integrate information in their working memory.

### Table 7. The mean value comparisons of the four conditions

<table>
<thead>
<tr>
<th>Group</th>
<th>PG</th>
<th>PSG</th>
<th>PTG</th>
<th>PTSG</th>
<th>High</th>
<th>Low</th>
<th>Post hoc tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention score</td>
<td>11.5</td>
<td>11.7</td>
<td>10.59</td>
<td>8.48</td>
<td>11.23</td>
<td>9.83</td>
<td>PG &gt; PTG, PSG &gt; PTG, High &gt; Low</td>
</tr>
<tr>
<td>Transfer score</td>
<td>5.67</td>
<td>5.13</td>
<td>4.74</td>
<td>6.78</td>
<td>7.65</td>
<td>6.43</td>
<td>PSG &gt; PG</td>
</tr>
<tr>
<td>Total score</td>
<td>17.17</td>
<td>19.83</td>
<td>18.07</td>
<td>15.26</td>
<td>18.88</td>
<td>16.26</td>
<td>PSG &gt; PTG, High &gt; Low</td>
</tr>
<tr>
<td>Intrinsic load</td>
<td>5.79</td>
<td>5.26</td>
<td>4.81</td>
<td>5.85</td>
<td>5.02</td>
<td>5.79</td>
<td>Low &gt; High</td>
</tr>
<tr>
<td>Extraneous load</td>
<td>5.67</td>
<td>4.96</td>
<td>4.85</td>
<td>5.81</td>
<td>5.02</td>
<td>5.6</td>
<td>PTSG &gt; PSG</td>
</tr>
<tr>
<td>Germane load</td>
<td>6.04</td>
<td>4.96</td>
<td>5.41</td>
<td>6.22</td>
<td>5.33</td>
<td>5.98</td>
<td>PG &gt; PSG</td>
</tr>
<tr>
<td>Retention load</td>
<td>7.83</td>
<td>5.83</td>
<td>5.67</td>
<td>6.37</td>
<td>6.27</td>
<td>6.53</td>
<td>PG &gt; PTG, PTG &gt; PTSG, PG &gt; PSG</td>
</tr>
<tr>
<td>Transfer load</td>
<td>8.29</td>
<td>6.39</td>
<td>6.93</td>
<td>7.15</td>
<td>7.08</td>
<td>7.28</td>
<td>PG &gt; PSG</td>
</tr>
<tr>
<td>Total load</td>
<td>33.63</td>
<td>27.39</td>
<td>27.67</td>
<td>31.41</td>
<td>28.73</td>
<td>31.19</td>
<td>PG &gt; PS</td>
</tr>
<tr>
<td>Study time</td>
<td>255.75</td>
<td>176.13</td>
<td>408.37</td>
<td>168.3</td>
<td>259.69</td>
<td>250.83</td>
<td>PTG &gt; PTG</td>
</tr>
<tr>
<td>Clicks</td>
<td>81.67</td>
<td>22.78</td>
<td>71.11</td>
<td>70.22</td>
<td>62</td>
<td>62.72</td>
<td>PG &gt; PSG</td>
</tr>
</tbody>
</table>

However, simultaneous presence of both cueing and verbal text caused interference was possibly because simultaneously presenting multi-faceted information in the visualizations without specificity distracted the learners from the focus (Imhof et al., 2013; Moreno, 2007). Perhaps the learners’ attention was captured more by
either salient cueing or verbal text than by surrounding information thereby preventing them from effectively gaining a conceptual understanding and mental model constructions (Kriz & Hegarty, 2007). Secondly, probably the presence of cueing that contrasted conspicuously with surrounding information preferentially attracted the learners’ attention; however, the verbal text that did not compete with the cueing for the learners’ attention but subservient to it (Boucheix et al., 2013), in which conceptual understanding was impaired. Thirdly, probably the cueing was detected but failed to continuously direct the learners’ attention across extended viewing of the visualization; therefore, the effects of cueing was compromised (Lowe & Boucheix, 2011). Finally, small screen size may have been another obstacle (Kim & Kim, 2012), since the text, pictures, and cueing were densely compacted on the small screens. Visual scanning to locate information on small screens could have likely proven to be too challenging for the students to perceive and comprehend the visualizations (Kim & Kim, 2012). Scanning information on the small screens made the students feel uncomfortable (Stockwell, 2012; Thornton & Houser, 2005), because it could easily affect their thinking processes and reasoning skills, while imposing a cognitive load on them (Kim & Kim, 2012). In addition, personal factors, such as the learners’ visuospatial ability (Boucheix & Lowe, 2010; Hegarty et al., 2003), cue habituation (Lowe & Boucheix, 2011), motivation, or metacognitive strategies (Kriz & Hegarty, 2007) could also diminish the impact of visuospatial cueing.

Research question 2

Those who received no visual cueing to reduce their visual search processes consumed their cognitive resources on processing non-essential information in the visualization. The cost of failing to extract highly relevant information from the visualization was that the quality of mental model construction would be compromised (Boucheix et al., 2013), thereby deep learning did not take place and yielded a high cognitive load. The presence of visual cueing that captured the learners’ attention might reduce the learners’ attention demands in selection of relevant information. With the reduction of attention demands, more cognitive resources would be released for essential information processing involved in the organization or integration of information (Jamet et al., 2008) and reduced the learners’ cognitive load. The presence of verbal text also helped the learners make referential connections between text and images and reduced their cognitive load. However, simultaneous presence of both cueing and verbal text yielded similar cognitive load as those who received nothing was probably the learners applied cognitive resources to process either the conspicuous cueing or verbal text and missed the essential information. If that was the case, it did not help them organize information into a coherent structure, fail to optimize conceptual understanding and impose a heavier cognitive load on them. The findings were partially in line with the study of Kalyuga et al. (1999) but were inconsistent with those of previous studies (e.g., Crooks et al., 2012; de Koning et al., 2010a; de Koning et al., 2011a; Lin & Atkinson, 2011) in which the presence of attention cueing yielded similar cognitive load.

Research question 3

Those who received attention cueing generally spent less time learning the lesson than their counterparts. This result only partially echoed the findings in Lin and Atkinson’s (2011) study but was inconsistent with the findings in Boucheix and Lowe’s (2010) study. Cue engagement may not guarantee cue obedience (Boucheix et al., 2013), and cue obedience may not guarantee cue consequence (Lowe & Boucheix, 2011). More cue loyalty and fast cue engagement may not guarantee better comprehension (Boucheix et al., 2013). Although visual cueing was successful in perceptually directing learners’ attention, the higher-order cognitive processes may not occur (Boucheix et al., 2013). Those in the PG received no visual cueing to reduce their visual search process, they were elicited to click more often and had longer viewing duration than their counterparts. Or perhaps the terminology in the instructional visualizations was too difficult, prompting those in the PTG to click more and had longer viewing duration than their counterparts. Those in the PTSG had short viewing duration when each time they clicked on the visualization was probably because the conspicuous cueing preferentially attracted their attention and triggered them to click frequently; however, the neighboring but highly-relevant information failed to compete equally with conspicuous cues for their attention (Boucheix et al., 2013), in which conceptual understanding was impaired. Besides, short viewing duration and high frequency of clicks implied more exploratory processing, such as visual searching and scanning (Lowe & Boucheix, 2016). Consuming cognitive resources on unnecessary visual searching was helpless for deep processing, whereas long viewing duration and low frequency of clicks might involve deep learning. Another possible explanation was that simultaneously presenting multiple-information on a small screen was more cognitively demanding to cause eyestrain (Kim & Kim, 2012; Stockwell, 2008), and prompting them to finish the task quickly or even give up without spending sufficient time carefully studying the visualizations. Besides, a lack of motivation (Liu & Chu, 2010) or other
personal factors might also prevent the learners from completing the activities on mobile phones (Stockwell, 2008) thereby compromising the effects of visual cues.

**Conclusion**

The experimental results indicated that the presence of either visual cueing or verbal text was beneficial for learners to develop greater conceptual understanding. However, if cueing failed to continuously attract learners’ attention across extended viewing of the visualization, the effects of cueing would be diminished. Whether the presence of progressive path cueing (Boucheix et al., 2013) or spreading-color cueing (Boucheix & Lowe, 2010) can provide learners with continuous attraction of attention by establishing pathway of causal chains and result in favorable mental model development is suggested to be probed in the future. Secondly, whether the verbal text presented in salient color initially and then switched to its original color can compete equally with conspicuous arrow cueing for learners’ attention is also suggested to be explored. Thirdly, relation cueing or anti-cueing (Lowe & Boucheix, 2011) is also suggested to be provided. With the diminishing of cue obedience after initial cue engagement, learners should shift their attention to neighboring but relevant areas in order to build a coherent, organized, and high-quality mental model (Boucheix et al., 2013). Finally, visual cueing that initially captures learners’ attention might wane over time due to learners’ cueing habituation or metacognitive strategies (Lowe & Boucheix, 2011), and these personal factors deserve further exploration.

**References**


Effects of Self-explanation and Game-reward on Sixth Graders’ Algebra Variable Learning

Hong-Zheng Sun-Lin* and Guey-Fa Chiou

Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taipei City, Taiwan // hzsunlin@gmail.com // gueyfa@ntnu.edu.tw

*Corresponding author

(Submitted March 15, 2016; Revised May 17, 2016; Accepted June 15, 2016)

ABSTRACT

This study examined the interaction effects of self-explanation and game-reward strategies on sixth graders’ algebra variable learning achievement, learning attitude, and meta-cognitive awareness. A learning system was developed to support the learning activity, and a 2×2 quasi-experiment was conducted. Ninety-seven students were invited to participate in this experimental instruction and assigned to four groups: self-explanation with game-reward, self-explanation, game-reward, and control. The results showed that (1) a significant interaction effect was found on the students’ learning achievement of algebra variable: the self-explanation with game-reward group performed significantly better than the self-explanation group, and both the game-reward group and self-explanation group scored significantly higher than the control group; (2) a significant interaction effect was found on the students’ learning attitude toward algebra variable: the self-explanation with game-reward group reported significantly more positively than the self-explanation group, and the control group responded significantly more positive results than the self-explanation group; (3) no significant interaction effect was found on the students’ meta-cognitive awareness of algebra variable learning: while the game-reward did not show a significantly positive effect, the self-explanation did.

Keywords

Self-explanation, Game-reward, Algebra variable learning, Learning attitude, Meta-cognitive awareness

Introduction

Algebra variable learning

Many students struggle with algebra variable concepts (Novotná & Hoch, 2008; Warren, 2003). MacGregor and Stacey (1997) noted that students usually do not understand how to use variables, the basic and essential concepts of algebra, to help them solve problems accurately and efficiently. Thus, Stacey and MacGregor (1997) suggested that students could be better prepared for further algebra learning by frequently using algebra notation and symbolism (i.e., the variable system) in learning contexts. According to the suggestion, Ross and Willson (2012) conducted an instructional experiment and found that engaging students in the given content with practical examples led to the better understanding of algebra variables. However, to date, issues regarding students’ algebra variable learning, such as use of certain strategies and design of examples, are still only partially explored (Bush & Karp, 2013).

As to proper understanding of algebra variables (i.e., the symbol system), Lucariello, Tine, and Ganley (2014) pointed out some key concepts. First, a symbol must be interpreted as representing an unknown quantity (Küchemann, 1978; McNeil et al., 2010), meaning a student must realise that a symbol represents a unit that does not have a certain value. Second, a student must interpret a symbol as representing a varying quantity or range of unspecified values (Kieran, 1992; Philipp, 1992). This concept is named as the multiple values interpretation of literal symbols (Knuth, Alibali, McNeil, Weinberg, & Stephens, 2005; Lucariello et al., 2014). To explore learners’ understanding of these two concepts, Knuth et al. (2005) tested seventh and eighth graders with a problem: “The following question is about the expression ‘2n+3.’ What does the symbol (n) stand for?” The correct responses should represent ideas that (1) the symbol could represent an unknown value (e.g., “the symbol is a variable, and it could stand for any value”) and (2) the symbol could represent more than one value (e.g., “it could be 3, 74, or even 123.4567”). The results, however, showed that approximately 39% of seventh graders and 25% of eighth graders answered incorrectly. The researchers thus suggested that an appropriate training would be necessary before middle-school education (e.g., when in sixth grade) for students’ better understanding of variables.

The third concept involves the relationships between symbols, and in such relationship the values change in a certain approach (e.g., as X increases, Y decreases, Lucariello et al., 2014). For example, Küchemann (1978) invited and tested 3000 high-school students who had learnt variables with the following problem: “Which is the
larger, 2n or n+2? Explain.” The results revealed that just only 6% of the students correctly realised the concept that the relation between 2n and n+2 is actually changing with n. Knuth et al. (2005) also explored this understanding of algebra variables by presenting middle-school students with similar problems. Likewise, the results showed that only approximately 18% of sixth graders, 50% of seventh graders, and 60% of eighth graders were aware of the relationship between symbols because their values change systematically.

Self-explanation strategy

Self-explanation, a concept-oriented learning activity, is defined as generating explanations to oneself to make sense of new and known information to be correct, and it has been regarded as an effective strategy in students’ algebra learning (Chi, 2000; Chi, de Leeuw, Chiu, & Lavancher, 1994; Neuman, Leibowitz, & Schwarz, 2000). For example, Aleven and Koedinger (2002) facilitated students’ self-explanation by using intelligent instructional software for algebra, and they found that students who explained their steps of problem-solving outperformed those who did not. According to Roy and Chi (2005), the process of self-explanation can encourage four forms of cognition: (1) recognise what information is missing while generating inferences, (2) integrate information into what learnt from a lesson, (3) relate new information to learners’ prior knowledge, and (4) identify and correct information received. This means that self-explanation strategy not only immerses students in cognitive processes but also engage them in the meta-cognitive activities to monitor what learning content they feel confused about.

While positive impacts of self-explanation have been consistently identified on producing robust learning gains in algebra learning (Aleven & Koedinger, 2002; Chi et al., 1994; Neuman et al., 2000), one problem still remains: the meta-cognitive activity could pose a challenge to middle-school students, such as sixth and seventh graders (e.g., Hausmann & Chi, 2002; Kuhn & Katz, 2009; Nokes, Hausmann, VanLehn, & Gershman, 2011). In self-explanation, students are usually required to deal with new and known information via question answering and answer explaining that could be challenging to be accomplished in a period of time (De Koning, Tabbers, Rikers, & Paas, 2011). Griffin, Wiley, and Theide (2008) encouraged students to self-explain through providing them with reading material and similar questions to practise, and they found that self-explanation could facilitate students’ better learning performance and deeper thinking. However, they also noted that if students were not required to record their self-explanations in any forms, there was no significant difference between self-explanation and non-self-explanation groups. This was supported by McEldoon, Durkin, and Rittle-Johnson (2013). In their study, students were encouraged to generate their explanations via rephrasing and writing to deepen their understanding of learning content, and the results indicated positive effects on learning achievement.

Another issue is the method of recording students’ self-explanation because it could affect the representation and structure of explanations (Hausmann & Chi, 2002). For example, generating explanations via typing (or writing) facilitate students’ monitoring and reflection on learning progress because it enables them to keep records and modify their explanations. Such method could be more appropriate for learners to employ than verbal ones that are generally less filtered (Chou & Liang, 2009). These typed explanations would encourage learners to properly self-monitor their comprehension, boost their reflective activity, and raise their meta-cognitive awareness (Griffin et al., 2008; Nokes et al., 2011). In the current study, therefore, all the participants were guided to type their explanations by the system, and they could review and modify these records during the learning activity. While above-mentioned studies pointed out the design issue of self-explanation activity and suggested the possible solution that facilitate student to self-explain with less challenges and obstacles, to date, few studies have been done on this issue.

Game-reward strategy

Gamification is the application of game elements to motivate target’s involvement in specific contexts such as business and education (Deterding, Dixon, Nacke, O’Hara, & Sicart, 2011; Landers & Callan, 2011; McGonigal, 2011; Sun-Lin & Chiou, 2017). By employing game-reward in teaching and learning, students could be encouraged to participate in tedious tasks with raised learning interest (Glover, 2013). However, Hanus and Fox (2015) noted that the concept of gamification encompasses many different game elements and their applications, thus it is difficult to distinguish and explore every possible facet of gamification. They attempted to apply game-reward strategy by providing students with a leader-board, coins, and badges; the results, however, showed that the non-game-reward group outperformed game-reward group. They mainly attributed such results to gamification forms and whether students were interested in the learning materials. They thus suggested that
gamification research should investigate specific elements/types first, such as points, coins, or badges, rather than an overarching concept so that learning effects can be taken more effectively.

The game-reward strategy takes only a small part of game design and can be used to benefit students’ learning achievement and learning attitude (Dicheva, Dichev, Agre, & Angelova, 2015; Hattie & Timperley, 2007). Nicholson (2015) noted that a game-reward learning system could be designed around the concept of “point” to represent the basic unit of exchange. Such exchange shows that how well and how much players earn for taking on the specific behaviour. The underlying concept of game-reward is simple: offering points to motivate players (Dicheva et al., 2015; Hanus & Fox, 2015). Nicholson (2015) stated that motivating players through points is similar to motivating people through other forms of incentives such as money or grades. From perspective of learning game design, points provide students with feedback and encourage their certain behaviour. Therefore, game-reward can be used to promote several learning activities, for example, motivating students to read given material, answer questions, explain answers, and review concepts.

Although game-reward is often employed in business and marketing, its educational applications are partially explored. Hamari, Koivisto, and Sarsa (2014) conducted a comprehensive review of empirical studies of gamification across different contexts (e.g., education and consumer science), but only 24 studies were identified. The authors noted methodological problems such as a lack of empirical evidence with the studies, and of the 24 reviewed studies, few actually compared the effects of gamified and non-gamified strategies in terms of students’ learning experience. Sun-Lin and Chiou (2017) attempted to examine the effects of gamified tasks on students’ algebra learning, and positive results were reported. Although some benefits of gamification such as game-reward have been found on students’ mathematics learning, empirical studies are still scant.

The combination of self-explanation and game-reward

According to Roy and Chi (2005), self-explanation could help to produce average learning gains via texts (22%), diagrams (44%), and multimedia presentation (20%). However, there are insufficient empirical evidence of game element effects with self-explanation. Johnson and Mayer (2010) ever reported that adult students who received self-explanation prompts in a learning game outperformed those who did not. Likewise, Hsu, Tsai, and Wang (2012, 2014) found that when students engaged in responding to prompts with game-reward, fourth graders’ science concept learning was significantly improved. Despite positive findings reported by these studies, a problem remains: learning-within-gaming may burden students with many different sub-activities or sub-processes (e.g., play games and answer questions) in a period of time.

O’Neil et al. (2014) conducted an experiment in which they added self-explanation prompts requiring players to answer one of three questions after completing each level of a math game regarding addition of fractions. They found that the requirement of self-explanation might result in extraneous cognitive process that distracted students’ attention and slowed their learning progress. Another issue was also proposed that students might respond quickly in order to return to the gaming section, but at the same time, they probably missed the opportunities of deeper thinking (e.g., reflection via self-explanation). Their findings indicated possible interaction between self-explanation and game elements, and a better combination should not interrupt the coherence of students’ learning and gaming process. Game-reward, as a form of gamification, provides students with incentives to focus their attention on learning tasks without interrupting the coherence, might be a possible solution. As to the design of self-explanation, Johnson and Mayer (2010) and Mayer and Johnson (2010) stated that explanation prompts could be useful as responding prompt might be easier for students to generate self-explanations than explaining without any guidance. They also mentioned that such design of prompts and learning tasks should be carefully considered because they may still require high cognitive effort that usually affect students’ engagement with learning content. While some positive impacts of self-explanation and gamification were reported respectively, little research has been done on their interaction effects on beginners’ algebra variable learning.

Research questions

In this study, a learning system was designed to support the combination of self-explanation and game-reward. A 2x2 quasi-experiment was conducted on a mathematics course of a primary school to explore the interaction effects on students’ algebra variable learning via investigating the following research questions:

- What are the effects of self-explanation and game-reward strategies, and the interaction between them, on students’ learning achievement of algebra variable?
• What are the effects of self-explanation and game-reward strategies, and the interaction between them, on students’ learning attitude toward algebra variable?
• What are the effects of self-explanation and game-reward strategies, and the interaction between them, on students’ meta-cognitive awareness of algebra variable learning?

Method

Participants

The participants of this instructional experiment included four classes of sixth graders, who were in the 11-13 age range ($M = 11.93$, $SD = .06$). A total of 97 students (47 females and 50 males) participated in this instructional experiment, each one class was assigned to a group: S-G group ($n = 24$), S-NG group ($n = 24$), NS-G group ($n = 24$), and NS-NG group ($n = 25$) (see Table 1).

<table>
<thead>
<tr>
<th>Table 1. The treatments (learning tasks) for the four groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game-reward (G)</td>
</tr>
<tr>
<td>Self-explanation (S)</td>
</tr>
<tr>
<td>Non-self-explanation (NS)</td>
</tr>
<tr>
<td>The NS-G group: non-self-explanation with game-reward</td>
</tr>
</tbody>
</table>

The learning system

A system, which combined self-explanation and game-reward strategies to support students’ algebra variable learning, was developed via software Construct 2. It provided nine learning tasks for students, and each unit comprised (1) a reading page introducing an algebra concept (Figure 1), (2) a sample question and the solution demonstrating the application of the concept (Figure 2), and (3) a question for practice helping students reflect on what they learnt (Figure 3).

Figure 1. A screenshot of algebra concept introduction

Figure 2. A sample question and the solution

The self-explanation strategy in this study required the participants to rephrase the algebra concepts they learnt, and to explain their answers to practice questions and each step of their solutions. The learning system was to
represent the entire learning procedure and materials, including reading material that introduced given algebra concepts, sample questions and the solutions that helped students understand how to apply the concepts, and practice questions that required students to answer and explain their answers with each step of solutions.

The game-reward strategy in this study was to reward students by points. These points were given according to their self-explanations of algebra concepts after answering questions (e.g., rephrasing concepts, which they just learnt, by using their own words to identify all steps of their solutions). Correct answers and logical explanations would earn them points that could be used to enhance their aircrafts in a mini game.

**Figure 3. A practice question**

![A question for students’ practice](image)

![Students explain their own answer and each step of the solution by typing](image)

![Students can go back to review if necessary](image)

**Instruments**

In this study, two algebra achievement tests (pre-test and post-test), an algebra learning attitude scale (pre-survey and post-survey), and an algebra meta-cognitive awareness questionnaire (pre-survey and post-survey) were used. The purpose of the pre-test/pre-surveys was to understand whether the four groups had an equivalent prior knowledge of algebra variables, prior learning attitude, and prior meta-cognitive awareness of algebra variable learning before attending the learning activity. Such results would be used as the covariate to employ appropriate statistics techniques if they were unequal between groups.

**Algebra variable achievement test**

The achievement tests were modified from an algebraic assessment developed by Lucariallo, Tine, and Ganley (2014). The pre-test and the post-test both consisted of nine multiple-choice items that covered three concepts of algebra variable: (1) the variable must be interpreted as representing an unknown quantity (three items), (2) students must interpret the symbols as representing a varying quantity (three items), and (3) the relationship exists between symbols as their values change in a systematic manner (three items).

To ensure the quality of measurement, 133 sixth graders, who had not participated in the instructional experiment, were recruited to answer the test questions. The reliability of the tests had been estimated based on Cronbach’s $\alpha$ measure for the pre-test $\alpha = .855$ and for post-test $\alpha = .939$; there was a strong positive correlation between the pre-test and post-test ($r = .397$, $p = .000$). In addition, the tests were reviewed and revised by two elementary-school mathematics teachers with more than five years of teaching experience, and this suggested the content validity. The tests also showed significant correlations with the regular school mathematics exam, which involved algebra variable concepts, and the significant positive correlations (between the exam and the pre-test, $r = .350$, $p = .000$; between the exam and the post-test, $r = .214$, $p = .010$) indicated that the tests had good criterion-related validity.

**Algebra learning attitude scale**

The scale regarding algebra learning attitude, which covered students’ enjoyment, motivation, self-confidence, and perceived value, was modified from the inventory developed by Lim and Chapman (2013). It comprises 19 items using a four-point Likert rating schema. Examples: “I enjoy learning algebra concepts”, “I am willing to learn more algebra concepts”, and “I am confident that I could learn algebra well.”
To ensure appropriateness of the questionnaire, the invited teachers helped to review and revise all the items. To examine the reliability, the 133 sixth graders were also invited to respond every item of the scale. The Cronbach’s $\alpha$ value for the entire scale was .927, showing acceptable reliability in internal consistency.

Algebra meta-cognitive awareness questionnaire

The meta-cognitive awareness questionnaire was modified from the instrument created by Panaoura, Philipou, and Christou (2003). It consisted of 27 questions that were developed based on a four-point Likert scale. Here are some examples of question items: “I know how well I have understood the algebra concepts I have learnt”, “I know what makes me find it difficult to solve algebra problems”, and “I examine my own performance while learning new algebra concepts.”

Like previous instruments, the two elementary-school teachers reviewed and revised all the questions, and the 133 students were invited to answer the questionnaire. The Cronbach’s $\alpha$ value for entire questionnaire was .959, showing acceptable reliability in internal consistency.

Experimental procedure

The experimental instruction took about 320 minutes in four weeks. First, the researchers explained purpose of the instruction (10 minutes), and all participants were required to answer the pre-test of algebra variable concepts (25 minutes) and responded the pre-surveys of learning attitude scale (15 minutes) and the meta-cognitive awareness questionnaire (15 minutes).

During the learning activity, the students learnt algebra variable concepts and accomplished nine learning tasks in 200 minutes. The differences of learning strategies between groups were the tasks that combined self-explanation (or non-self-explanation) with game-reward (non-game-reward).

After the learning activity, all groups answered the post-test of algebra variable concepts (25 minutes), and responded post-surveys of learning attitude scale (15 minutes) and the meta-cognitive awareness questionnaire (15 minutes).

Results

Learning achievement

The assumption of homogeneity was tested by Levene’s test ($F = .991, p = .401$), and the two-way ANCOVA results (see Table 2) indicated a significant interaction effect found between the game-reward and self-explanation ($F = 8.312, p = .005, \eta^2 = .083$) on students’ learning achievement of algebra variable concepts. It was necessary to conduct a simple main effect analysis to investigate specific learning effects. The descriptive data of students’ adjusted test scores for all groups are shown in Table 3.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-explanation</td>
<td>.280</td>
<td>1</td>
<td>.280</td>
<td>.097</td>
<td>.001</td>
</tr>
<tr>
<td>Game-reward</td>
<td>4.265</td>
<td>1</td>
<td>4.265</td>
<td>1.485</td>
<td>.016</td>
</tr>
<tr>
<td>Self-explanation*game-reward</td>
<td>23.882</td>
<td>1</td>
<td>23.882</td>
<td>8.312$^*$</td>
<td>.083</td>
</tr>
<tr>
<td>Error</td>
<td>264.328</td>
<td>92</td>
<td>2.873</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $^*$p < .05.

As to the results grouped by self-explanation strategy, firstly, a significant difference ($F = 4.197, p = .048, \eta^2 = .102$) was found between the S-G group the S-NG group, as shown in Table 4. The learning achievement of the S-G group (adjusted $M = 5.397$) was significantly greater than that of the S-NG group (adjusted $M = 4.797$). According to the principles of effect size proposed by Cohen (1988), the partial eta squared ($\eta^2$) of the results of the simple main effect analysis represented a moderate effect ($\eta^2 = .102 > .059$).

Secondly, Table 4 presents that a significant difference ($F = 8.340, p = .006, \eta^2 = .130$) was found between the NS-G group and the NS-NG group. To be specific, learning achievement of the NS-G group (adjusted $M =$
5.968) was significantly higher than that of the NS-NG group (adjusted $M = 4.497$), as shown in Table 3. The partial eta squared represented a moderate effect ($\eta^2 = .130 > .059$).

<table>
<thead>
<tr>
<th>Table 3. The descriptive data of each group’s learning achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-explanation strategy</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Self-explanation</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Non-self-explanation</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. The simple main effect analysis results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pair</strong></td>
</tr>
<tr>
<td>S-G versus S-NG</td>
</tr>
<tr>
<td>NS-G versus NS-NG</td>
</tr>
<tr>
<td>S-G versus NS-G</td>
</tr>
<tr>
<td>S-NG versus NS-NG</td>
</tr>
</tbody>
</table>

Note. $p < .05$.

As to the results grouped by game-reward strategy (see Table 4), there was a significant difference between the S-NG group and the NS-NG group ($F = 21.010, p = .000, \eta^2 = .296$). The former gained significantly higher score than the latter, and the partial eta squared indicated a strong effect ($\eta^2 = .296 > .138$). However, no significant difference of the learning achievement was found between the S-G group and the NS-G group ($F = .135, p = .716$).

**Learning attitude**

After verifying the assumption of homogeneity of regression ($F = .997, p = .398$), the two-way ANCOVA results (see Table 5) indicated an interaction effect of the strategies ($F = 9.549, p = .003, \eta^2 = .094$) on students’ learning attitude toward algebra variable. It was thus necessary to conduct a simple effect main effect analysis. The descriptive data of the adjusted students’ survey scores are shown in Table 6.

<table>
<thead>
<tr>
<th>Table 5. The two-way ANCOVA results of learning attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>Self-explanation</td>
</tr>
<tr>
<td>Game-reward</td>
</tr>
<tr>
<td>Self-explanation*game-reward</td>
</tr>
<tr>
<td>Error</td>
</tr>
</tbody>
</table>

Note. *$p < .05$.

As for the grouped results by self-explanation strategy, a significant difference ($F = 12.083, p = .001, \eta^2 = .251$), as Table 7 presents, was found between the S-G group and the S-NG group. The learning attitude of the S-G group (adjusted $M = 61.904$) was significantly more positive than that of the S-NG group (adjusted $M = 46.990$). The partial eta squared of the results indicated a strong effect ($\eta^2 = .251 > .138$). However, no significant difference was found between the NS-G group and the NS-NG group ($F = .595, p = .444$).

<table>
<thead>
<tr>
<th>Table 6. The descriptive data of each group’s learning attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-explanation strategy</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Self-explanation</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Non-self-explanation</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 7, in terms of grouping results by game-reward strategy, although no significant difference ($F = 1.431, p = .238$) was found between the S-G group and the NS-G group, but the NS-NG group (adjusted $M = 49.468$) reported significantly more positively ($F = 6.107, p = .017, \eta^2 = .111$) than the S-NG group (adjusted $M = 46.990$). The partial eta squared represented a moderate effect ($\eta^2 = .111 > .059$).
Meta-cognitive awareness

With the verification that the assumption of homogeneity of regression was not violated through Levene’s test ($F = 2.624, p = .055$), the two-way ANCOVA result was shown in Table 8. The interaction effect between the game-reward and self-explanation was not significant ($F = .349, p = .556$). Thus, it was sensible to directly examine the main effects of the independent variables. The results of main effects analysis, as shown in Table 9, indicated that significant effects were confirmed for self-explanation strategy ($F = 5.465, p = .022, \eta^2 = .059$) but not confirmed for game-reward strategy ($F = .159, p = .691, \eta^2 = .002$). This implies that the post-survey score of students were significantly different due to self-explanation strategy, as described in Table 8. To be specific, students who learnt with self-explanation strategy (the S-G and the S-NG groups; adjusted $M = 81.867$) reported significantly higher meta-cognitive awareness of algebra variable learning than those who learnt with non-self-explanation strategy (the NS-G and the NS-NG groups; adjusted $M = 73.795$). The partial eta squared of the results of self-explanation strategy indicated a moderate effect.

### Table 7. The simple main effect analysis results

<table>
<thead>
<tr>
<th>Pair</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>$F$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-G versus S-NG</td>
<td>1522.885</td>
<td>1</td>
<td>1522.885</td>
<td>12.083</td>
<td>.251</td>
</tr>
<tr>
<td>NS-G versus NS-NG</td>
<td>128.075</td>
<td>1</td>
<td>128.075</td>
<td>.595</td>
<td>.011</td>
</tr>
<tr>
<td>S-G versus NS-G</td>
<td>280.289</td>
<td>1</td>
<td>280.289</td>
<td>1.431</td>
<td>.033</td>
</tr>
<tr>
<td>S-NG versus NS-NG</td>
<td>1087.568</td>
<td>1</td>
<td>1087.589</td>
<td>6.107</td>
<td>.111</td>
</tr>
</tbody>
</table>

Note. $p < .05$.

### Table 8. The two-way ANCOVA results of meta-cognitive awareness

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>$F$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-explanation</td>
<td>1445.483</td>
<td>1</td>
<td>1445.483</td>
<td>5.465</td>
<td>.059</td>
</tr>
<tr>
<td>Game-reward</td>
<td>41.994</td>
<td>1</td>
<td>41.994</td>
<td>.159</td>
<td>.002</td>
</tr>
<tr>
<td>Self-explanation*game-reward</td>
<td>92.256</td>
<td>1</td>
<td>92.256</td>
<td>.349</td>
<td>.004</td>
</tr>
<tr>
<td>Error</td>
<td>23013.310</td>
<td>92</td>
<td>264.521</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $p < .05$.

### Table 9. The descriptive data of each group’s meta-cognitive awareness

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Adjusted $M$</th>
<th>SE</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-explanation</td>
<td>81.867</td>
<td>2.681</td>
<td>24</td>
</tr>
<tr>
<td>Non-self-explanation</td>
<td>73.795</td>
<td>2.246</td>
<td>24</td>
</tr>
<tr>
<td>Game-reward</td>
<td>77.129</td>
<td>2.452</td>
<td>24</td>
</tr>
<tr>
<td>Non-game-reward</td>
<td>78.532</td>
<td>2.475</td>
<td>25</td>
</tr>
</tbody>
</table>

**Discussion and conclusions**

Overall, this study explored the interaction effect between self-explanation and game-reward on sixth graders’ algebra variable learning, and a 2×2 quasi-experiment was conducted. The experimental results showed that (1) a significant interaction effect was found on students’ learning achievement: the S-G group performed significantly better than the S-NG group, and the NS-G group and the S-NG group gained significantly higher scores than the NS-NG group; (2) a significant interaction effect was found on students’ learning attitude: the S-G group reported significantly more positive results than the S-NG group, and the NS-NG group responded significantly more positively than the S-NG group; (3) no significant interaction effect was found on students’ meta-cognitive awareness: self-explanation strategy significantly affected students’ responses, but game-reward strategy did not.

**A significant interaction effect on algebra variable learning achievement**

First, the S-G group performed significantly better than the S-NG group, and second, the NS-G group scored significantly higher than the NS-NG group. These results showed that students could be encouraged to learn given concepts by receiving points as rewards. Although students were not required to explain their answers, they were willing to read the instructional materials and try to understand the algebra variable concepts to correctly answer questions. This means that although game-reward (i.e., earn more points for game-play), as an
extrinsic motivator, does not directly relate to learning content, it could still motivate students to learn and gain higher scores than those who do not receive game-reward. These results are in accord with the views of Dicheva et al. (2015) and Nicholson (2015). The game-reward strategy can work to promote several learning activities, including reading learning material, answering questions, and reflecting on their performance. The results of this study could also address issues proposed by O’Neil et al. (2014) and Hanus and Fox (2015). O’Neil et al. (2014) noted that self-explanation would interact with game elements and thought the combination should not interrupt the coherence of students’ learning and gaming; Hanus and Fox (2015) stated that gamification research should investigate specific elements, such as points or badges, rather than an overarching design so that the learning effects can be effectively taken. In this current study, the benefits of game-reward strategy were found to encourage students to learn without distracting them from learning process, and thus their learning achievement would be improved more effectively.

Third, the S-NG group gained significantly higher test scores than the NS-NG group. The result revealed that supporting students to explain their answers would have a positive effect on their learning achievement. Students can deepen their understanding of given concepts through the process of self-explanation. They read the instructional material and rephrase it via their own expressions, that is, from input to output in the learning process. Such process would promote students to reflect on what and how they have learnt, and it also provides them with the access to re-read the material when they find it difficult to answer questions and generate explanations. Through this process they could improve the understanding of algebra variable concepts. This result supports findings of Griffin et al. (2008) that self-explanation activity, which are implemented via providing reading instructions and related examples for practice, could benefit students’ learning achievement. This result also evidences the ways of explanation generation mentioned by McEldoon et al. (2013). They suggested that promoting students to generate their explanations in a certain manner, such as writing or typing, might effectively help them relate self-explanations to instructional material. This study found that typing explanations helped students keep records and provide them with an opportunity to easily revise, and such approach seems feasible to be used in class.

A significant interaction effect on algebra variable learning attitude

Firstly, the S-G group reported significantly more positively than the S-NG group. The experimental results revealed that game-reward strategy could enhance both students’ learning attitude and their learning achievement rather than divert their attention. This is consistent with many previous studies which reported positive effects of game-reward on learning attitude (e.g., Glover, 2013; Hattie & Timperley, 2007). This result also provides further exploration for Attali and Arieli-Attali’s (2015) findings that point manipulation showed only minor effects on students’ learning performance. The points, used to show students’ accuracy in their study, probably lack following applications of the reward, that is, the usage or function of these points. In this current study, points that students gained via accomplishing learning tasks connected to game-play. Students could use these points to enhance attributes of their roles (i.e., aircrafts) in a mini game. According to the result, such game-reward successfully encouraged students to participate in and accomplish learning tasks, and this seems to be a feasible application of gamification for teaching.

Secondly, the NS-NG group responded significantly more positive results than the S-NG group. This indicated that self-explanation did not significantly improve students’ attitude toward algebra variable learning. The result supports the finding of Kuhn and Katz (2009). It is possible that students could learn better via self-explanation, but they might find it difficult and make more effort to accomplish learning tasks. They may thus be under some pressure that results in negative learning attitude. This is probably because without appropriate incentives (e.g., game-reward or other rewards), students might feel bored and be unwilling to stick to learning especially when they are required to accomplish challenging tasks (e.g., generate explanations), even if such activity could help them understand the learning material better. A possible explanation may lie in students’ abilities in self-explanation. They found it difficult to generate explanations probably because of a lack of mastered self-explanation skill. This could be a threshold that students need to cross. If they have an opportunity to practise self-explanation until they could generate self-explanations without too many difficulties, they would be able to get a sense of achievement in the learning process, and their learning attitude might be more positive than that in this study.
Self-explanation enhanced meta-cognitive awareness of algebra variable learning

Although no interaction effect between self-explanation and game-reward strategies was found, self-explanation showed a significantly positive impact on students’ meta-cognitive awareness of algebra variable learning. It not only provided students with opportunities to deepen what they learnt but help them raise their meta-cognitive awareness. By explaining the answer with students’ own words, they could reflect on what and how much content they could understand. The results are consistent with the findings of several previous studies revealing that guiding students to explain answers with their own expressions can boost meta-cognitive awareness (Kuhn & Katz, 2009; Roy & Chi, 2005).

Moreover, the results may also address a proposed issue that the meta-cognitive requirements of self-explanation lead to higher difficulty for middle-school students (e.g., Hausmann & Chi, 2002; Nokes et al., 2011). In this study, there was a similar practice question following each example in the system, and when students found it difficult to answer questions or generate explanations, they could go back to previous pages to re-read the instructional material without any limitation. The practice question was to encourage students to test themselves, and the access to demonstration pages was to deepen their understanding through review if necessary. The design may reduce the difficulty of accomplishing learning tasks and lead students to reflect on their learning performance and learning process.

References


135


The Influence of a Pedagogical Agent on Learners’ Cognitive Load

Noah L. Schroeder
Wright State University, Dayton, OH, USA // noah.schroeder@wright.edu

(Submitted March 24, 2016; Revised July 9, 2016; Accepted August 10, 2016)

ABSTRACT
According to cognitive load theorists, the incorporation of extraneous features, such as pedagogical agents, into the learning environment can introduce extraneous cognitive load and thus interfere with learning outcome scores. In this study, the influence of a pedagogical agent’s presence in an instructional video was compared to a video that did not contain a pedagogical agent. The results indicated no significant differences on a multiple choice learning outcome measure or on a perceived mental effort measure that was rated both after the instructional sequence and after completion of the post-test. In addition, no significant differences were found between groups in relation to their training efficiency or instructional efficiency. Accordingly, the position that the mere appearance of a pedagogical agent in the learning environment increases the cognitive load imposed by a learning task should be re-examined.

Keywords
Pedagogical agent, Virtual human, Cognitive load, Mental effort

Introduction
Virtual characters continue to appear in a multitude of computer-based learning environments. One type of virtual character, the pedagogical agent, continues to be a focus of recent research. The term pedagogical agent refers to a wide range of virtual characters, but they always have a visual presence in the learning environment and their purpose is to help students learn the material (Moreno, 2005; Park, 2015; Yung & Paas, 2015).

While pedagogical agent research has been ongoing for approximately 20 years, researchers have continued to question the efficacy of a pedagogical agent’s ability to facilitate the learning process. While Heidig and Clarebout’s (2011) systematic review found that pedagogical agents often led to no significant differences between groups on learning outcomes, Schroeder, Adesope, and Gilbert’s (2013) meta-analysis found that pedagogical agents can lead to small positive effects on learning outcomes. Despite these findings, the author concurs with Heidig and Clarebout (2011) in that “the question of whether pedagogical agents generally facilitate the learning process is too broad” (p. 30). However, it is important to note that this statement could apply to nearly any type of technology-based learning tool. As such, it continues to be important to examine what types of outcomes various pedagogical agents can influence in the learning environment.

Proponents of pedagogical agents have argued that agents can enhance the social aspects of the multimedia learning environment, thus causing the learner to be more cognitively engaged with the learning materials (Atkinson, Mayer, & Merrill, 2005; Mayer & DaPra, 2012; Mayer, Sobko, & Mautone, 2003). Similarly, researchers have found evidence that the pedagogical agent’s design can influence learning outcomes and learners’ perceptions (Domagk, 2010; Veletsianos, 2007). However, we still do not have definitive guidance as to what features pedagogical agents should have and there are many possible agent implementations yet to be thoroughly investigated.

As with many other technological innovations, the use of pedagogical agents is somewhat controversial. Arguments against their use are typically grounded in cognitive load theory. For instance, Clark and Choi (2007) argued that pedagogical agents may overload a learner’s working memory capacity. This notion is echoed in Wouters, Paas, and van Merriënboer’s (2008) review, which concluded that agents must be designed carefully to avoid overloading the working memory capacity. Similarly, van Mulken, André, and Müller (1998) argued that a pedagogical agent may cause distraction to the learner during the learning process.

Another issue that deserves further attention in the realm of multimedia learning is the pacing of instruction. Most pedagogical agent research has been conducted in learner-paced environments, where the instruction is broken into pre-determined instructional segments (Schroeder, Adesope, & Gilbert, 2013). However, developing this sort of program often requires computer programming skills and time that many instructors do not have. Accordingly, many instructors will host videos on services like YouTube. A YouTube video shares qualities of both learner-paced and system-paced instructional sequences. For instance, the learner can pause and forward – qualities that are similar to a learner-paced environment. However, the instructional sequence is not
segmented and does not pause and wait for learners to click to continue – qualities that are similar to a system-paced instructional sequence. Since some principles of multimedia learning can have differential, or even inverse effects in differently paced environments (Ginns, 2005), it is important to clarify the type of pacing being used. Schroeder and Adesope (2013, 2015) have identified the pacing that occurs in these types of videos as learner-attenuated system-paced (LASP) learning environments. Despite the fact that LASP learning environments are very common in educational settings, limited work exists examining how multimedia learning principles influence learning or perceptive outcomes when learning with pedagogical agents in these environments. To the author’s knowledge, only two studies have examined the use of pedagogical agents in a LASP learning environment (Schroeder & Adesope, 2013, 2015), and neither considered the impact on learners’ cognitive load.

The purpose of this study is to explore the influence of a pedagogical agent on learners’ cognitive load in a LASP learning environment. In the next section, perceived cognitive load, pedagogical agents, and the efficiency of instructional sequences are discussed in relation to their theoretical underpinnings and connected with recent research.

**Literature review**

**Cognitive load theory**

According to theorists, human cognition relies on three essential and interacting components: the working memory, the long-term memory, and organizational structures called schema (Paas & Sweller, 2014; Sweller, 2005). In Paas and Sweller’s (2014) view, the long-term memory acts as an information repository or archive. This archive is organized by structures called schema, which are essentially series of interconnected mental models (Sweller, 2005). However, neither of these structures process new information. New information is processed through the working memory, which can only handle a finite amount of information at once (Paas & Sweller, 2014; Sweller, 2005). While a limited processor sounds like it would interfere with learning, Paas and Sweller (2014) suggested that the limitations actually make it possible for us to process all the information around us in our daily lives.

Cognitive load theorists have described three types of cognitive load a learner may experience during a learning task. These are known as intrinsic cognitive load, extraneous cognitive load, and germane cognitive load. Intrinsic cognitive load is caused by the level of complication of the learning material itself, which means that it varies for each individual learner (Kalyuga, 2011; Paas & Sweller, 2014; Sweller, 2005, 2010). The more complex the information is, and the more pieces or chunks of information that must be processed at one time in order to understand the content, the more intrinsic cognitive load the material causes (Sweller, 2010). However, poorly designed instructional sequences can also cause learners to exert more mental effort to learn the material than should have been required, from processes such as complicated visual searches for salient information or content which is presented physically distant from relevant images (Kalyuga, 2011; Paas & Sweller, 2014; Sweller, 2005, 2010). Theorists have termed the mental effort required for these tasks as being due to extraneous cognitive load. Finally, germane cognitive load is due to the mental effort required to actually integrate the new information into meaningful knowledge structures for that learner (Kalyuga, 2011; Paas & Sweller, 2014; Sweller, 2005, 2010). For a visual representation of how the above concepts may interact during a learning task see Paas, Tuovinen, Tabbers, and Van Gerven (2003, see Figure 1).

For many years researchers have created measures to examine what features within educational environments may cause intrinsic, extraneous, or germane cognitive load. However, many attempts to create such instruments have been criticized, and scholars have posited it may not be possible to measure each type of cognitive load individually (Kirschner, Ayers, & Chandler, 2011; Sweller, 2010). However, the mental effort scale (Paas, 1992) has been used for decades without many objections (Kirschner et al., 2011).

Paas’s (1992) mental effort scale is a one-item measure of perceived mental effort and it has been broadly used as an indication of the cognitive load a learner experienced. The measure has also been used to calculate how efficient an instructional condition is (see Paas & van Merriënboer, 1993). However, Paas et al. (2003) noted that researchers had not been using the measure of cognitive load in the efficiency calculation as it was initially intended. Paas and van Merriënboer (1993) intended for mental effort to be measured after the problem solving phase of the research study, which leads to a calculation known as instructional efficiency. Paas et al. (2003) found that many researchers were instead using standardized scores of mental effort ratings measured after the instructional sequence, but before problem solving. This calculation, while nearly identical mathematically, leads to a different outcome Paas et al. (2003) classified as training efficiency. While the nuance between the two types
of efficiency seems minor – stemming from when mental effort was measured – it impacts the interpretation of the results. While instructional efficiency refers to the mental effort required to solve the problems, training efficiency refers to the mental effort required to learn the material from the intervention. Fundamentally these are similar, yet different measures. Until recently, neither efficiency measure had been considered in relation to pedagogical agent research.

Previous research with pedagogical agents and cognitive load measures

Few pedagogical agent researchers have examined the influence of virtual characters on learners’ cognitive load outcomes compared to non-agent conditions (see Schroeder & Adesope, 2014). Most that have examined cognitive load outcomes used measures such as perceived ease (Frechette & Moreno, 2010) or perceived difficulty (Atkinson, 2002; Moreno, Mayer, Spires, & Lester, 2001). Fewer yet have calculated the training efficiency or instructional efficiency of the intervention. The author is aware of only two studies in which researchers have examined some form of efficiency when learning with pedagogical agents compared to non-agent conditions.

Choi and Clark (2006) examined the cognitive efficiency of learning with a pedagogical agent compared to a non-agent condition. However, it is important to note that they used a different formula than Paas and van Merriënboer (1993). In one condition, learners worked with a pedagogical agent that signaled their attention to appropriate parts of the screen. In the control condition, the learners had the same learning environment, except arrows replaced the virtual character. Their results showed that there were no significant differences on the cognitive efficiency outcomes between the two groups.

More recently, Yung and Paas (2015) compared the use of a pedagogical agent to a non-agent condition when teaching students about certain aspects of the circulatory system. The agent used visual cues, such as pointing, to help guide the learners’ attention to relevant items on the screen. In their analysis, they examined instructional efficiency using Paas and van Merriënboer’s (1993) formula. Their analysis of the results indicated that learners in the pedagogical agent condition had a significantly better relationship between their self-reported mental effort and their performance scores, meaning that the environment that contained a pedagogical agent was more efficient than the non-agent condition (Yung & Paas, 2015).

As shown, limited research examines the efficiency of pedagogical agents, as well as their impact on cognitive load outcomes in general. Similar to findings around learning outcomes, research around the cognitive load implications of including a pedagogical agent in the learning environment has produced mixed results (Schroeder & Adesope, 2014). Thus, more work is needed to understand to what extent pedagogical agents influence cognitive load outcomes and the training and instructional efficiencies of an intervention.

Research questions

It is clear that more systematic research is needed to understand how, to what extent, and under what circumstances pedagogical agents can influence learning outcomes. Similarly, it is apparent that we know little about what the inclusion of a pedagogical agent means in relation to cognitive load outcomes, yet concerns about the introduction of extraneous cognitive load are a primary reason why agents are often cited as unimportant, or even potentially ineffective, in the learning environment.

Due to the aforementioned gaps in the extant literature, this study sought to examine the influence of a pedagogical agent in a LASP learning environment through the investigation of the following research questions:

- **RQ1** - How does the inclusion of a pedagogical agent influence learners’ cognitive outcomes compared to a non-agent condition?
- **RQ2** - How does the inclusion of a pedagogical agent influence learners’ self-reported mental effort compared to a non-agent condition?
- **RQ3** - How does the inclusion of a pedagogical agent influence learners’ training efficiency compared to a non-agent condition?
- **RQ4** - How does the inclusion of a pedagogical agent influence learners’ instructional efficiency compared to a non-agent condition?
Methods

Participants

The participants in this study were 75 pre-service teachers at a public university in the United States. Six participants’ scores were removed due to missing data. Hence, 69 participants’ scores were examined. The average age of the participants was 21 years old ($SD = 4.105$) and 84% were female. The participants were low prior knowledge learners, with 92.8% reporting that they’d never received any formal instruction around multimedia learning theory. Furthermore, no participants scored any points on the pre-test ($M_{overall} = 0.00$, $SD_{overall} = 0.00$). The participants were randomly assigned to either the control group ($n = 33$) or the experimental group ($n = 36$).

Computer-based materials and design

The learning materials for this study consisted of a 486 word narrative presented by a recorded human male voice. The narrative was a slightly expanded version of the narrative used in Schroeder and Adesope (2013, 2015). The background consisted of six slides designed to provide a low-verbal redundancy learning environment. In other words, the main ideas were summarized on each slide in typical bullet point design, but the text did not align verbatim with the spoken narration.

![Figure 1. The control condition](image1.png)

![Figure 2. The virtual human (pedagogical agent) condition](image2.png)
The pedagogical agent-based learning environment was created using a combination of Microsoft Powerpoint, Media Semantics Character Builder software by Media Semantics Inc. (see http://mediasemantics.com/), and Camtasia Studio 8 by TechSmith Corporation (https://www.techsmith.com/camtasia.html). In the control condition (Figure 1), learners viewed a LASP video that was essentially a voice-over Powerpoint presentation. In the experimental condition (Figure 2), the virtual human remained stationary in the lower-right hand corner of the screen. In order to reduce confounding variables, the agent did not move or gesture, however it did have its lips synchronized to the narrative to create the illusion of a more life-like character. The survey was implemented using Qualtrics, an online survey software. The LASP instructional video was hosted on YouTube and then embedded into the Qualtrics survey.

Pre-test

The pre-test consisted of the same three free-response knowledge questions as in Schroeder and Adesope (2013, 2015). The questions were, “State six things you know about cognitive load theory,” “Describe the split-attention principle,” and “Describe the modality principle.” Scoring followed the same procedures as Schroeder and Adesope (2013, 2015), and there was a maximum score of 14 points available. For the first question regarding cognitive load theory, points were issued for correctly identifying (1 point) or describing (1 point) germane, intrinsic, or extraneous cognitive load, as well as the long-term memory, schema, or working memory (Schroeder & Adesope, 2013, 2015). The second and third questions were each worth one point, which was earned by correctly describing the principle in question (Schroeder & Adesope, 2013, 2015).

Post-tests

There were three post-tests involved in this study. The first measure was the mental effort rating scale developed by Paas (1992). In this case, the wording of the item was slightly modified to be more appropriate to having just completed viewing an instructional video. The item was stated as, “In studying the preceding video I invested” and was answered through a 9-point Likert scale ranging from very, very low mental effort through very, very high mental effort.

The second measure was a 30-item multiple choice test. The measure was very similar to the instrument used in Schroeder and Adesope (2013, 2015), and measured the major concepts presented in the narrative. In an effort to establish rudimentary validity evidence for the measure, it was reviewed by a faculty member familiar with the concepts presented in the script. The internal consistency reliability was found to be $\alpha = .75$.

The final measure was Paas’s (1992) original mental effort instrument. The item was answered through a 9-point Likert scale ranging from very, very low mental effort through very, very high mental effort.

Procedure

The study took place in a classroom that contained one desktop computer for each student. Headphones were provided to the participants if necessary. Participation in the study was optional and learners were not compensated for participating. Overall, participating took approximately 30 minutes.

Participants first completed the demographic questions, followed by the pretest. After the pretest, they either watched an instructional video containing a pedagogical agent or one that did not, depending upon the experimental condition they were assigned to. After viewing the video, the participants answered the modified version of Paas’s (1992) mental effort scale, followed by the multiple choice measure, and then finally Paas’s (1992) original mental effort scale. The researcher was present throughout the duration of the data collection.

Results and discussion

In this section, the results of the study are addressed and discussed in relation to each research question and previous research.
RQ 1 – How does the inclusion of a pedagogical agent influence learners’ cognitive outcomes compared to a non-agent condition?

An independent t-test was conducted to examine if any differences existed between groups’ learning outcome scores. Levene’s test for equality of variances was significant $F = 4.520$ ($p = .037$), and thus equal variances were not assumed. The results of the independent t-test showed no significant differences between groups $t(66.14) = .021$ ($p = .983$). As shown in Table 1, the means of both groups were nearly identical.

Table 1. Means and standard deviations of learning outcomes for each group

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>15.64</td>
<td>4.51</td>
</tr>
<tr>
<td>Experimental</td>
<td>15.61</td>
<td>5.53</td>
</tr>
</tbody>
</table>

The finding of no significant differences between groups on learning outcomes is consistent with some other studies investigating the effectiveness of a pedagogical agent compared to a non-agent condition (Heidig & Clarebout, 2011). In this case, the result was somewhat expected due to the fact that the agent did not do anything to specifically facilitate the learning process. Rather, in order to minimize confounding variables, the pedagogical agent just stood next to the content and provided narration. Thus, these results show support of the “Persona Zero-Effect” (Miksatko, Kipp, & Kipp, 2010, p. 3) in that the pedagogical agent’s presence did not significantly influence learning outcomes.

RQ 2 – How does the inclusion of a pedagogical agent influence learners’ self-reported mental effort compared to a non-agent condition?

As mentioned, the learners’ self-reported mental effort was measured at two separate points in time in order to differentiate training efficiency from instructional efficiency. To examine this research question, the raw scores of the mental effort measures themselves are examined.

In relation to the self-reported mental effort scores measured directly after watching the instructional video, the examination of the data indicated that it was not normally distributed. Squareroot, log, and reciprocal transformations failed to improve the normality of the distribution. Accordingly, non-parametric statistics were used to analyze the data. Despite the fact that the experimental group’s mean score reflects that learning from the instructional video required slightly less mental effort than the control group (Table 2), the results of a Mann-Whitney test revealed that there were no significant differences between the control group ($Mdn = 5.00$) and the experimental group’s ($Mdn = 5.00$) scores, $U = 541.50$, $z = -.656$, $p = .512$, $r = -.079$.

Table 2. Means and standard deviations of perceived mental effort for each group immediately after the instructional video

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>4.85</td>
<td>1.20</td>
</tr>
<tr>
<td>Experimental</td>
<td>4.69</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Self-reported mental effort scores were also examined from when the learners finished the multiple choice test. Preliminary analysis of the data indicated that the data were normally distributed. Hence, an independent samples t-test was used to test for any significant differences between groups. Levene’s test showed equal variances could be assumed $F = .014$ ($p = .907$). Analysis of the results showed that while the experimental group reported a higher mean mental effort required to solve the problems (Table 3), this effect was not statistically significant ($t(67) = -1.404$, $p = .165$).

Table 3. Means and standard deviations of perceived mental effort for each group immediately after problem solving

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>5.27</td>
<td>1.57</td>
</tr>
<tr>
<td>Experimental</td>
<td>5.81</td>
<td>1.58</td>
</tr>
</tbody>
</table>

While at first glance non-significant results may seem unimportant, in this case they are actually quite enlightening. As previously mentioned, arguments against the incorporation of pedagogical agents generally hinge on the notion that pedagogical agents have the potential to increase the extraneous cognitive load present within the learning environment (Clark & Choi, 2007; van Mulken, André, and Müller, 1998; Wouters, Paas, &...
van Merriënboer, 2008). Presumably, the concern is that the agent is a distraction from the learning materials (van Mulken, André, & Müller, 1998). In this study the agent was not designed to facilitate learning in any specific way through gesture, rather it was merely present throughout the video. Hence, if agents were a source of distraction to the extent some seem to imply, it is plausible this would have been reflected in the mental effort scores. It must be stated however, that the agent used in this study did not use gestures, move across the screen, or otherwise try and cause a distraction from the learning content. Thus, the results of this study show that the mere incorporation of a pedagogical agent may not necessarily increase the cognitive load imposed by a LASP learning environment, but the author reiterates Wouters et al.’s (2008) comment that agents should be carefully designed. In the future researchers should explore the impact of an agent that signals, gestures, or otherwise moves or interacts with the learner (i.e., the agent should do more than simply provide the narrative) to examine the impact on cognitive load outcomes. However, in such a study it would be important to minimize confounding variables between conditions.

RQ 3 – How does the inclusion of a pedagogical agent influence learners’ training efficiency compared to a non-agent condition?

First, the standardized scores were calculated from both the self-reported mental effort after watching the instructional video as well as the multiple choice test. The scores were then entered into Paas and van Merriënboer’s (1993) formula.

An independent samples t-test was used to compare the two groups’ scores (Table 4). Levene’s test showed that equal variances could be assumed ($F = 2.91$, $p = .093$). The results showed that there were no significant differences between groups ($t(67) = -.331, p = .742$).

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>-.04</td>
<td>.81</td>
</tr>
<tr>
<td>Experimental</td>
<td>.04</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Table 4. Means and standard deviations of each groups’ training efficiency

To the author’s knowledge, no other studies investigating pedagogical agents compared to non-agent conditions have calculated training efficiency, so it is difficult to draw many conclusions based on one study. However, since the differences between groups did not reach statistical significance, it is safe to say that the mere incorporation of a pedagogical agent into the LASP learning environment did not hinder nor benefit the learners’ training efficiency. Reconceived, including a pedagogical agent may not necessarily make the content easier for the students to learn from a LASP instructional video. Thus, as advised by other scholars, pedagogical agents should be carefully designed (Veletsianos, Miller, & Doering, 2009; Wouters et al., 2008), yet more research needs to be conducted to examine how different features agents may possess, such as signaling to relevant information, influence the training efficiency.

RQ 4 – How does the inclusion of a pedagogical agent influence learners’ instructional efficiency compared to a non-agent condition?

First, the standardized scores were calculated from both the self-reported mental effort after completing the multiple choice test as well as the multiple choice test itself. The scores were then entered into Paas and van Merriënboer’s (1993) formula.

An independent $t$-test was used to examine if differences existed between the two groups (Table 5). Levene’s test showed that equal variances could be assumed ($F = .203$, $p = .654$). The results showed that there were no significant differences between groups ($t(67) = 1.041, p = .301$).

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>.13</td>
<td>.94</td>
</tr>
<tr>
<td>Experimental</td>
<td>-.12</td>
<td>.97</td>
</tr>
</tbody>
</table>

Table 5. Means and standard deviations of each groups’ instructional efficiency

As mentioned, to the author’s knowledge only one study has examined the instructional efficiency of a pedagogical agent compared to a non-agent condition. The results of this study are contrary to Yung and Paas’s (2015) findings. Yung and Paas (2015) found that a pedagogical agent significantly improved the instructional
efficiency of the environment compared to the non-agent condition. Presumably, this is due to the difference in agent features between the studies. In this study, the agent was relatively static, while in Yung and Paas’s (2015) study the agent signaled the learner’s attention through gestures. Since the signaling principle and its benefits for learning and reducing extraneous cognitive load are relatively well established (Mayer & Fiorella, 2014), it should not be surprising that the signaling agent in Yung and Paas’s (2015) study had a more favorable instructional efficiency than the non-agent condition. A useful next step would be to conduct a study comparing a static agent to an agent that gestures to examine if the gestures used to signal content are what influence the instructional efficiency, or if the difference in results is attributable to some other difference between the studies.

Conclusion

This study marks a preliminary attempt to comprehensively examine the cognitive load implications of including a pedagogical agent into a LASP learning environment. The results of this study have implications for both theory and practice.

In relation to cognitive load theory, this study adds to the literature around training efficiency and instructional efficiency. As mentioned, these measures are rarely used in the context of pedagogical agents. The results of the study show that the incorporation of a relatively static agent did not significantly influence cognitive load outcomes, training efficiency, or instructional efficiency. Thus, an agent does not necessarily make the content easier to learn, nor make the problem solving process any more difficult. However, more work is needed to examine the different features pedagogical agents may possess, such as signaling, movement, and the ability to demonstrate tasks in order to examine if they have any influence on cognitive load outcomes or measures of efficiency.

For those considering implementing an agent, the results of this study indicate that the mere incorporation of an agent does not necessarily increase the cognitive load of a LASP learning environment. Hence, choosing to incorporate an agent will not necessarily impede student learning. Rather, evidence from recent reviews suggests that the incorporation of an agent will either not make a significant difference to learning outcomes (Heidig & Clarebout, 2011) or may even facilitate learning outcomes (Schroeder et al., 2013). In fact, Moreno (2005) stated that no studies had found that a pedagogical agent impeded learning compared to a non-agent condition. Despite these findings, the author reiterates calls for the thoughtful design of pedagogical agents (Wouters et al., 2008, Veletsianos et al., 2009). As Veletsianos et al. (2009) stated, “pedagogical agent integration in educational settings should be guided by the added-value opportunities that agents present for enhancing the social, pedagogical, and technological opportunities provided to learners” (p. 179). Hence, while we have not yet uncovered many specific features that help agents to be widely effective in different content domains, these features may become apparent if researchers continue to systematically investigate agents that add value to the learning environment rather than merely being present.

In closing, the findings of this study must be interpreted with the understanding of the inherent limitations of its design. First, the study used a type of pacing rarely examined in pedagogical agent research. Thus, a similar study should be conducted in a learner-paced environment to see if the results are transferrable. Similarly, the content of the learning sequence was conceptual in nature rather than procedural. Due to this, similar studies should be conducted with learning materials that require procedural knowledge to see if any significant differences are found. In addition, the instructional sequence was relatively short in duration. Future research should be conducted with learning materials that require a longer period of student interaction with the content and the pedagogical agent. Finally, this study did not examine the influence of a well-designed pedagogical agent meant to specifically facilitate learning in any meaningful way. In short, the agent was present, but did not gesture to important content or otherwise signal the learners’ attention. Being that increasing the pedagogical affordance of the learning environment is a key notion behind pedagogical agent implementation, it would be salient to examine the influence of a well-designed pedagogical agent compared to a static agent. This question was outside the scope of this study, but would provide important considerations for pedagogical agent researchers.

References


Development and Usability Test of an e-Learning Tool for Engineering Graduates to Develop Academic Writing in English: A Case Study

Chih-Chung Lin¹, Gi-Zen Liu²* and Tzong-I Wang²

¹Department of Foreign Languages & Literature, National Cheng Kung University, Tainan City, Taiwan // ²Department of Engineering Science, National Cheng Kung University, Tainan City, Taiwan // ul8701459@gmail.com // gizen@mail.ncku.edu.tw // wt535@mail.ncku.edu.tw

*Corresponding author

(Submitted July 19, 2016; Revised December 14, 2016; Accepted February 23, 2017)

ABSTRACT

Many non-native English speaking (NNES) graduates are required to write academic papers in English; consequently, recent research in the past decade has been devoted to investigating the usefulness of genre-based Writing Instructions (GBWI) on learners’ writing cultivation. There is little specific guidance, however, on how GBWI can be employed in online learning environments. The study aimed to design an e-learning tool with a Chinese-based interface and to evaluate its perceived usefulness on users’ English academic writing development. Needs analysis and GBWI were employed as the main approaches for the system development, for which 32 participants involved in the trail. The results indicate that the participants considered the system useful and helpful while learning English academic writing and extract the design principles derived from the empirical results in the study. The findings provide both learning content developers and e-learning tool designers with developmental insights for graduates to acquire academic writing ability in English. In conclusion, this study contributes to the understanding of GBWI acting on the NNES engineering graduates’ development of English academic writing knowledge and skills with the development of e-learning tools.

Keywords

Intelligent tutoring system, Engineering English, Scientific writing, Usability test

Introduction

In the recent years, globalization has placed increasing attention on the issue of using English as a primary medium in academia (Barrett & Liu, 2016; Glasman-Deal, 2009). English has now become a major language for people around the world, and is also used across various disciplines in the academic industries as the dominant language used for indexed journal publication (Curry & Lillis, 2010; Huang, 2010; Lee & Lee, 2013; Liu, Lo, & Wang, 2013; Liu, Lin, Kou & Wang, 2016). For global competitiveness, international rankings, and deeper involvement in academic development, publishing journal articles has become an essential part and indispensable pursuit for all institutions in higher education, particularly for professors and young scholars such as doctoral students who are striving to publish their academic works in indexed journals all over the world. While all seem to agree that academic writing ability in English plays a vital role across disciplines in academic industry, problems and challenges have been raised as to how researchers, young scholars, or even graduate students can cope with this trend. For instance, Berman and Cheng (2010) compared the perceived difficulties of academic writing in English between native speakers (NS) and non-native speakers of English students (NNES). The results showed that although both groups perceived the ability of academic writing in English as necessary, NNES in English as Foreign Language (EFL) context appeared to be more challenging than their NS counterparts. An investigation on the difficulties reported by Taiwanese researchers conducted by Min (2014) illustrated that EFL Taiwanese academics faced both linguistic and non-linguistic challenges such as grammar, tense, cohesion and logic of the overall passages when submitting their academic articles written in English to the journals.

Recently, a number of applied linguists have called for a reflection on how to help EFL graduate and undergraduate students cultivate their English academic writing proficiency (Liu, Lo, & Wang, 2013; Liu, Lin, Kou & Wang, 2016). Genre-based writing instructions (GBWI) is a widely used teaching and learning approach to help learners develop their writing ability (Huang, 2010; Liu et al., 2016), in which learners are given explicit instructions on the moves and structures embodied in different types of writing to meet different rhetorical purposes, as well as for understanding the power relation of different move functions (Wingate, 2012).

However, there is a noticeable absence of research projects dealing with how learners can benefit from using e-learning tools to cultivate EFL engineering graduate students’ academic writing ability in English via GBWI. As for the teaching and learning of English academic writing, the integration of technology into teaching pedagogy has now become a popular and important issue (Chen, Shih & Liu, 2015; Lin, Kang, Liu & Lin, 2016; Liu, Wu, 2017).
& Chen, 2013; Loncar, Barrett, & Liu, 2014; Spence & Liu, 2013; Stoddart, Chan & Liu, 2016). A review investigating the affordance of Learning Technology (LT) on science education conducted by Liu, Chiu, Lin and Barrett (2014) found that LT can help learners develop English academic writing ability more effectively and efficiently. Also, the use of Information and Communication Technology (ICT) in the development of writing was also found significantly useful and helpful (Genlott & Grönlund, 2013). To this end, several e-learning systems have been designed to help users develop English writing proficiency, such as MyAccess (Lee, 2008), iWriting (see http://www.iwriting.com.tw/), and Criterion (Burstein, Chodorow & Leacock, 2004).

However, most writing systems are not specifically designed for the purpose of academic writing, and few e-learning learning systems have been specifically developed for users whose native language is Mandarin Chinese. Although an e-learning tool exists, namely EJP-Write (Lo, Liu & Wang, 2014), which was designed for Taiwanese graduate students to develop academic writing ability in English across disciplines (Liu, Chiu, Lin & Barrett, 2014), Engineering English Journal Paper Writing, EEJP-Write, is specially designed for graduates whose native language is Mandarin Chinese in engineering-related disciplines to develop English academic writing ability. EEJP-Write combines many needed functions together. For example, EEJP-Write helps users arrange the affiliation of author(s), set a journal as the target one for submission, thus, helping them develop their papers in a more effective and efficient way by fully knowing the regulations and guidelines of that journal. EEJP can be adopted as an alternative teaching instrument for teachers and learners in the fields of Second Language Acquisition (SLA) and English for Specific Purposes (ESP) to reduce their workload and workload in teaching and learning L2 English academic writing. In particular, EEJP offers Taiwanese engineering graduates with a tailored e-learning tool to develop their academic writing proficiency in English.

The purpose of the present case study was to design and explore the perceived usefulness and usability of an e-learning tool incorporating GBWI. According to Thomas (2011), a case study is one of the important research methods in social science, and around 20% of the literature in social science done from 1975 to 2000 was conducted via this research method. The reason of using case study as the main research method was that there were few courses specifically designed for engineering graduate students to develop academic writing ability in English; however, publishing academic works in English was important and even compulsory especially for Ph.D students. Hence, the specific aims of this study include (a) to design an e-learning tool for Taiwanese engineering graduate students to develop academic writing ability in English, and (b) to elucidate some of the theoretical assumptions regarding the perceived usefulness and usability of the e-learning tool, Engineering English Journal Paper Writing, EEJP.

**Phase I: Design and development of EEJP**

**Needs analysis**

A Needs Analysis (NA) was conducted to ensure the overall design of courses or e-learning tools meets the needs of the target users (Chostelidou, 2011). Several studies have employed mixed research methods (see Dehnad et al., 2010; Spence & Liu, 2013), and the benefits suggest that different sources of data can empower researchers with a more complete understanding, thereby allowing for fuller and more informative explanations. Both quantitative and qualitative methods were used in the NA to support validity and reliability of the results from learners, teachers, and experts in the field of academic journal paper writing.

The NA results are presented in Table 1 and Table 2, and were compiled via triangulation among three groups, namely graduate students, thesis and dissertation advisors and department chairpersons. 124 graduate students, 32 theses and dissertation advisors, and 8 department chairmen volunteered to participate in the NA. As shown in Table 1, different departments have different criteria for graduate students. Some use the quantitative way through counting the number of published papers and the others employed qualitative way, in which graduate students can get the points based on the ranking of publications. All three groups reported that graduate students in Taiwan are required to publish at least one journal article as a graduation criterion, including writing theses and dissertations, conference papers, and indexed journal papers in English. Further, both doctoral students and department chairpersons suggested that graduate students publish two journal articles as part of their doctoral program. Table 2 lists the major difficulties and needed assistance of graduates while writing journal papers in English. How to meet the academic research writing standards in science-related fields, how to avoid plagiarism and how to organize and present ideas in English were mentioned by all three groups. Accordingly, the present researchers used these results as the source for the design and development of learning content and system functions in EEJP.
Table 1. Results of the needs analysis for publishing indexed journal papers in English as a graduation criterion

<table>
<thead>
<tr>
<th>Graduation criterion</th>
<th>Graduate students</th>
<th>Thesis/Dissertation advisors</th>
<th>Department chairmen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative standard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 paper</td>
<td>20.97%</td>
<td>4.35%</td>
<td>37.5%</td>
</tr>
<tr>
<td>2 papers</td>
<td>59.68%</td>
<td>4.35%</td>
<td>62.5%</td>
</tr>
<tr>
<td>3 papers</td>
<td>7.26%</td>
<td>8.70%</td>
<td>0.00%</td>
</tr>
<tr>
<td>4 papers</td>
<td>4.03%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Qualitative standard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 points</td>
<td>0.81%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>4 points</td>
<td>0.81%</td>
<td>4.35%</td>
<td>0.00%</td>
</tr>
<tr>
<td>5 points</td>
<td>7.26%</td>
<td>60.87%</td>
<td>0.00%</td>
</tr>
<tr>
<td>6 or more points</td>
<td>16.13%</td>
<td>17.39%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 2. Results of the major difficulties and needed assistance when writing journal papers in English

<table>
<thead>
<tr>
<th>Target goals of learning</th>
<th>Graduate students</th>
<th>Thesis/Dissertation advisors</th>
<th>Department chairmen</th>
</tr>
</thead>
<tbody>
<tr>
<td>How to meet the expected standard in the academic fields</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>How to cite others’ work and avoid plagiarism</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>How to organize and present ideas</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>How to systematically present the results/data</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>How to provide examples to support the ideas</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>How to clearly express their ideas via English while writing academic papers</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>How to use transition words in English academic writing</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>How to use appropriate tenses</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>How to present the same ideas with different words</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

System development

After the major needs and learning targets were identified through triangulation among the three groups, 10 engineering-related research fields were selected, including Material Science and Engineering, Biomedical Engineering, Electrical Engineering, Mechanical Engineering, Chemical Engineering, Computer Science and Information Engineering, Electronics Engineering, Photonics, Civil Engineering, and Medical Engineering. Instead of selecting the top Science Citation Index (SCI) journals, professors who had experience in instructing Ph. D students in each engineering-related department were asked to suggest 3 indexed journals as the major source to develop the learning content in EEJP. In each field, the three most-suggested journals were chosen as the major source for the development of the learning content in EEJP. In each journal, 5 articles were randomly downloaded and analyzed to identify the embodied moves and functions in the research reports. Thus, 150 journal articles in total were analyzed as the major learning content for EEJP.
All the journal papers were analyzed via genre analysis. After the moves, functions and overall structures were identified from the 150 selected articles, 10 professors in each field were invited to review the genres and moves and to write example sentences. The learning content and example sentences were then proofread by native speakers of English with experience in academic writing and engineering. Moreover, the IEEE citation style guide was also included because it offers advice for developing plagiarism-avoidance ability, as well as information for a better understanding of citation rules. The learning materials of IEEE citation style were based on the IEEE citation manual.

The overall architecture of the EEJP system is presented in Figure 1. Generally, there are three main modules in the EEJP, namely My Paper, References and Notes, and Introduction to Journal Paper Writing and Format. In My Paper, users can start their papers by adding the information such as the paper title, the author, the affiliation of the author, and the target journal or conference they want to submit to. In Reference and Notes, users can add new research topics to categorize the references, add related references, make notes from references, and import reference information from Endnote, which is a popular tool for arranging references. The last module, Introduction to Journal Paper Writing and Format, offers instructions on academic writing formats, writing templates, move structures, example sentences, tense and voice in academic writing, and IEEE citation style. The homepage of the EEJP system is presented in Figure 2.

![Figure 2. Mandarin Chinese interface of EEJP-Write](image1)

![Figure 3. The interface of “My Paper” in EEJP-Write](image2)
In *My paper*, users can initiate their academic writing by typing in the research titles, adding affiliation of the author(s), and setting information and notes of the target journals (such as word limits and special regulations). Further, they can also edit their papers in this area (see Figure 3). By inputting the above information, users create specific goals of what is required at the onset of the writing process. Also, in the overall process of writing academic papers, users may revise their papers anytime including the paper title, the content, tables, figures and reference notes.

After clicking on the paper title shown on the list on the FrontPage of *My Paper*, the EEJP system displays the online writing area (see Figure 4), which can be divided into three main parts. The left column shows the outline of the paper and the set information concerning the target journals. The writing area of EEJP system is in the middle, and is also where users can develop tables and figures, and write formulas. The buttons above the writing area contain common functions, such as typesetting including modifying font size, changing the aligning style and adjusting word pitch. The column on the right offers supportive tools for English writing, including an online bilingual dictionary with examples, synonyms and antonyms, reference notes, and a collocation checking system (see Figure 5).

![Figure 4. The online writing area interface in My Paper](image)

**Figure 4.** The online writing area interface in *My Paper*

![Figure 5. Functions of notes and references in the online writing area in My Paper](image)

**Figure 5.** Functions of notes and references in the online writing area in *My Paper*

When writing research papers in English, users may need to check the usage as well as the collocation of words. In the EEJP system, users can directly use the collocation dictionary embedded in the system (see Figure 6). The data of the collocation checking system is based on the journal papers, which comprise the main source of learning materials in EEJP, as mentioned. This function was included according to the NA, which suggested that users have difficulties in the collocations of English words.
The learning materials of EEJP are based on Swale’s (1990) genre approach, analyzing and categorizing the moves found in the selected journal papers. This enables users to follow the move structures and writing templates, and thereby acquire knowledge of writing academic papers in English. Each move in the writing templates includes the Chinese translation, meaning explanations, and example sentences to promote users’ understanding of what a move is and to guide them in the development of their academic works in English. The learning materials include the purposes and functions of each chapter in academic papers, move structure and writing templates, examples of each chapter in academic papers, tenses of each chapter in academic papers (see Figure 7).

**Phase II: Evaluation on EEJP**

The present study was designed to answer the following research questions: Are users satisfied with EEJP in terms of usefulness, usability, and personalization? We investigated this question by constructing and administering a questionnaire and semi-structured interviews. Three research questions are addressed in the following:

- In terms of the usefulness, does the learning content of EEJP help learners cultivate their academic writing ability in English? The considered criteria for evaluating the perceived usefulness of EEJP include: (a) relevance, (b) format, (c) reliability, (d) level and (e) timeliness.
In terms of the usability, do the system functions of EEJP help learners write their academic works in English more efficiently? The considered criteria for evaluating the usability of EEJP include: (a) ease of use, (b) aesthetic appearance, (c) navigation, (d) terminology, and (e) learnability.

What factors influence users’ perception toward EEJP? The considered criteria include the perceived usefulness, usability, personalization, attitude, and follow-up intention of using EEJP.

Questionnaire

A five-point Likert scale was employed in the questionnaire, in which the independent variables were participants’ demographics, including gender, educational level, writing anxiety, writing difficulty, and their prior experience of using e-learning tools. The dependent variables were their perceived usefulness, usability, personalization, attitude and follow-up intention of use toward EEJP.

The questionnaire consisted of seven main parts, namely personal information, writing and computer anxieties, usefulness, usability, system personalization, user attitude, and follow-up intention of use. The personal information part gathered the information of personal background, proficiency level, needs of writing in English, and prior experience of using e-learning tools. Writing anxiety and computer anxiety session investigated the degree of individual’s anxiety of English writing and of using computers. Usefulness part investigated the perceived content usefulness of the learning materials provided in EEJP system while usability explored the usability of the system functions. System personalization tried to know the perceived degree of personalization of the system and users’ attitude investigated users’ attitude after they used EEJP system. Last part, follow-up intention of use, was designed to know users’ intention of using EEJP system in the future. Moreover, Cronbach’s Alpha analysis was used to determine that the results of the survey questionnaire was reliable, and that each dimensions of the questionnaire was acceptable and correlated to prove the content validity and reliability of the quantitative data.

Semi-structured interviews

After completion of the survey questionnaire, some participants volunteered to participate in face-to-face interviews, which comprised seven semi-structured questions to gather more information regarding the details of their experience using EEJP. Questions considered usefulness, usability, personalization, attitude, and follow-up intention of using. Although the interview questions were based on the results of the survey questionnaire, the order of interview questions was not fixed, allowing interviewees to express their opinions, attitude, and suggestions in a natural and conversational way, while encouraging more information to be gathered (Mann, 2011). Each interview lasted about 25 to 35 minutes.

<table>
<thead>
<tr>
<th>Table 3. Cronbach’s Alpha of dimensions in the questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variables</td>
</tr>
<tr>
<td>Perceived usefulness</td>
</tr>
<tr>
<td>Perceived usability</td>
</tr>
<tr>
<td>Users’ attitude</td>
</tr>
<tr>
<td>Follow-intention of using EEJP-Write</td>
</tr>
</tbody>
</table>

The study employed a mix method approach to characterize user perceptions toward EEJP. In keeping with tradition in quantitative research, the validity and reliability of the questionnaire were examined through expert consulting and Cronbach’s Alpha analysis (see Table 3). Moreover, the semi-structured interviews were used to gather more information for identifying and confirming participants’ suggestions. Achieving this goal required triangulation of data sources. The study documented the participants’ thoughts and feelings of using EEJP over an extended period of time, revealing their experience, preferences and suggestions toward using an e-learning tool to develop their academic writing ability in English.

Sample

As mentioned, EEJP was designed for engineering English, and so graduate students from engineering-related disciplines were recruited. The sample of 32 participants (including 14 doctoral students and 18 master students) from the departments of engineering and computer science at a research-based university in southern Taiwan were recruited. The participants were native speakers of Mandarin, and expressed interests in learning and
improving their academic writing ability in English. Accordingly, participants were familiar with all functions of the EEJP system to facilitate their learning goals before filling in the questionnaire.

**Experimental procedure**

As mentioned above, all participants volunteered to participate in the research, and were required to attend a 1-hour introduction session prior to the trial of using the EEJP. During the session, the researchers explained the purpose of the overall research, and demonstrated how to use EEJP. Questions raised by the participants were answered in order to make sure that all the participants know how to use EEJP. After all participants were informed and fully aware of the goals and the tasks of the trial, they were then requested to use EEJP as though they were beginning to compose one academic paper, learning from the instructions presented in the learning materials provided in the system. The overall writing time lasted 4 hours. After the trial, participants were asked to fill out the questionnaire based on their experience of using EEJP system, which required about 20 to 30 minutes. Interviews were then conducted after the preliminary analysis of the questionnaire as a basis to explore the possible explanations.

**Data collection and data analysis**

The present study aimed at developing and evaluating an e-learning tool, EEJP, for graduate students whose native language is Mandarin Chinese in engineering departments to cultivate English journal paper writing ability. The first research question investigates users’ perceived usefulness of the learning materials provided in the EEJP system, including the instructions related to the genres and moves, instructions on verb tenses, IEEE citation guide, and move templates with example sentences. The second research question explores users’ perceived usability of the EEJP system in terms of interface design and functionality of the system to verify whether the system facilitates users writing academic papers. Third, the present study attempts to determine the possible factors under the framework of TAM, which would influence users’ attitude toward using EEJP system, and the fourth research question explores whether users’ attitude have direct and positive effect on the follow-up intention of using EEJP system.

The quantitative data were analyzed using SPSS version 17.0 statistical software, and presented via descriptive statistics. NVivo version 9.0 was used to analyze the qualitative interview data based on the concept-driven way to identify themes and topics. All responses were categorized into three dimensions in terms of usefulness, usability, and factors influencing follow-up intention of using.

**Results**

**Quantitative results**

Participants showed differences in the perceived usefulness and usability of EEJP. Table 4 illustrates the means and standard deviations for the four dimensions of the survey. It was found that participants held a positive attitude toward the learning content ($M = 4.18, SD = .64$), however, they also reported that the functionality of EEJP ($M = 2.96, SD = .78$) was not that useful. Nevertheless, participants still held a positive attitude ($M = 3.90, SD = .85$) and had the intention to use EEJP ($M = 3.96, SD = .59$). Table 5 presents the correlation results among the variables. As can be seen, there was a direct relation among perceived usefulness, perceived usability, user attitude and follow-up intention of using EEJP.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived usefulness</td>
<td>4.18</td>
<td>.64</td>
</tr>
<tr>
<td>Perceived usability</td>
<td>2.96</td>
<td>.78</td>
</tr>
<tr>
<td>Users’ attitude</td>
<td>3.90</td>
<td>.85</td>
</tr>
<tr>
<td>Follow-intention of using EEJP-Write</td>
<td>3.96</td>
<td>.59</td>
</tr>
</tbody>
</table>

Table 4. Survey results (n = 32)

Through applying paired-sample t-test, learners’ prior experience of using e-learning tools was found to influence users’ perceived usefulness and usability of LT. As shown in Table 6, results indicate that learner’s prior experience had a significant impact on users’ perception toward the learning content and functionality of EEJP and confirmed that users’ prior experience played a role in influencing users’ satisfaction regarding
learning contents and system functions in e-learning tools (Chen & Yao, 2016; Maillet, Mathieu & Sicotte, 2015). Further, different academic levels, be it master or doctoral students, did not influence users’ perception and attitude toward using EEJP.

<table>
<thead>
<tr>
<th>Table 5. Pearson Correlations among the variables in the survey questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
</tr>
<tr>
<td>Usefulness</td>
</tr>
<tr>
<td>Usability</td>
</tr>
<tr>
<td>Users’ attitude</td>
</tr>
<tr>
<td>Follow-up intention</td>
</tr>
</tbody>
</table>

Note. **p < .01.

<table>
<thead>
<tr>
<th>Table 6. Results of t-test among examined variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
</tr>
<tr>
<td>Education level</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Prior experience with e-learning tool</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Qualitative results

An overall summary of the synthesized results concerning the advantages, unsatisfactory items, and suggestions about the learning content are presented in Table 7. In terms of the usefulness of EEJP, the quantitative and qualitative data show that participants were satisfied with the EEJP learning content, while the writing structures and templates, as well as sentence examples and “Academic Organizations and formats” were reported to be useful and helpful for learning how to write academic works in English. In addition, the qualitative data suggest that users found the EEJP learning content useful and helpful (see Table 7) with respondents indicating that the content helps them develop academic writing ability in English.

<table>
<thead>
<tr>
<th>Table 7. Summary of the usefulness results in terms of advantages, unsatisfactory items, and suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unsatisfactory items</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Suggestions</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 8. Summary of usability results in terms of advantages, unsatisfactory items, and suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unsatisfactory items</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Suggestions</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
The results concerning the advantages, unsatisfactory items, and suggestions toward the system functions are presented in Table 8. As for the usability of EEJP, the quantitative results show that participants are satisfied with functions such as the online dictionary, the collocation dictionary, setting outlines and previewing the whole paper. Moreover, participants reported that such functions are what they need when writing academic works in English. However, some technical problems such as system halts and improper functioning were mentioned by the interviewees, which seriously influenced user satisfaction and attitude toward the EEJP system functions. Moreover, participants suggested that the operation of the “Online area” could be more similar to Microsoft Word, so that time learning how to use new functions would be reduced.

Derived design principles for GBWI learning content developers and e-learning tool designers

Based on both quantitative and qualitative results concerning the learning materials provided in EEJP, two design principles regarding the benefits of learning contents stated by the users in the study were extracted. They should provide some developmental insights for GBWI learning content developers to consider while designing the learning materials for EFL learners to cultivate English academic writing knowledge and skills of using e-learning tools.

Principle 1: Needs analysis and explicit approach of GBWI as interventions

Essential elements are figured out as the main sources to design and develop the learning materials through Needs Analysis (NA). In the study, NA was conducted as the first step serving as the main sources for the preparation of learning materials and users were found to be satisfied with the learning contents \( M = 4.18 \). Moreover, explicit approach of GBWI is considered useful and helpful for EFL learners while they are learning academic writing ability in English. In EEJP, users were given explicit instructions and they were guided through direct representations of writing templates. Following the clear guidance on the writing structures, users in the present study reported that the explicit presentation on rhetorical moves and genre structures were useful for them to develop academic writing ability in English.

Principle 2: Relevance and timeliness of the learning materials

Relevance and timeliness are crucial standards while preparing the learning materials with the focus of cultivating academic writing ability in English for EFL graduates. The purpose of the study is to design and develop an e-learning tool for EFL graduate students to develop English academic writing ability. 200 journal publications from the selected indexed journals published from 2012 to 2013 were chosen and analyzed as the main source of learning materials in EEJP, the journals were selected by the faculty who based their decisions on their experience of guiding graduates and on the relation to the field they study. Moreover, papers published from 2012 to 2013 in the selected journals were selected for genre analysis to reflect the real-time organizations of journal paper writing.

Principle 3: Navigation, terminology and button icons

Navigation and use of terminology are two practical issue while designing e-learning tools. Unlike other widely used writing learning systems such as MyAccess, iWriting, and Qbook, EEJP is one of the few tools which are specially designed for EFL graduates whose native language is Mandarin Chinese to learn how to write academic papers in English. It is relatively difficult for users to learn how to use EEJP effectively and efficiently within a short period of time. Thus, the navigation in EEJP was not that clear to users, thus, made it difficult to learn where and how to start using EEJP. Moreover, some terminology used in EEJP didn’t help users obtain the essence of that system function, which made it difficult for users to use some functions in EEJP such as the icon used for inserting a formula confused users. Thus, it is suggested that all the terminology and system icons be examined through a pilot study to ensure the legibility to users.

Principle 4: Users’ prior experiences as the references for designing system functions

Users’ prior experience and preferences are needed to be considered. When first using EEJP, most users in the present study tended to rely on their prior experience, and tried to use in the way they were familiar with. Users
self-reported that they had difficulty learning how to use EEJP since the operation of some system functions was quite different from Microsoft processor, Word. Thus, future e-learning tool designers can design the operation procedure in light of the way used in Microsoft processor, which might reduce users’ cognitive load and increase their follow-up intention.

Discussions

The present study employed GBWI as the main teaching and learning approach for the preparation of learning materials and for the design of learning activities embodied in EEJP, and users reported being highly satisfied with the learning materials. The overall responses concerning the perceived usefulness of EEJP indicate that users were satisfied with the EEJP learning content, which also reflects findings of previous GBWI studies regarding explicit instructions on the teaching and learning of journal paper writing in English for EFL graduate students (Bailey, 2010; Huang, 2010; Khodabandeh et al., 2013).

Genre-based Writing Instructions (GBWI) can be an effective teaching and learning approach for both teacher educators and learners. The quantitative results illustrate that the instructions on the rhetorical moves and structures along with sentence examples were regarded as the most useful part of EEJP. Based on the analyzed results, participants were found to be satisfied with the learning materials on EEJP (Mean > 3.5), and reported that the materials, including the writing templates, sentence examples, and instructions on rhetorical structures and functions as well as verb tenses, were satisfactory. One possible reason for this might be the presentation of GBWI in the preparation of the learning materials. GBWI has been found to effectively assist graduate students to develop academic writing competence in English (Englander, 2014; Lo et al., 2014). Lo et al. (2014) further reported that users perceived the English Journal Paper Writing Guide (EJP-Write) with the incorporation of GBWI as a useful and helpful e-learning tool for graduates in social science disciplines to develop academic writing ability in English. Another possible reason for its success may be that the design and development of the learning content were mainly based on the suggested criteria proposed by Tsakonas and Papatheodorou (2006) including “Relevance,” “Format,” “Reliability,” “Level,” and “Timeliness.” Taking these five criteria into consideration in the present study, the results concerning the perceived usefulness suggest that future learning systems should consider the same five criteria for the design and development of learning content for EFL graduate students to develop academic writing ability in English.

Both the quantitative and qualitative results suggest that the functions and features related to referencing, note-taking, the online dictionary, and the collocation dictionary were considered the most useful. However, some unsatisfactory items were revealed such as modifying the pitch interval, inserting tables and figures and making statistical formula. One reason that users may have felt the system functions might not be useful lie in the design of the systematic functions. Since the operation of some system functions is quite different from the way in which Microsoft Word operates, it may take some time for users to get familiar with some functions, such as developing mathematical formulas and inserting figures and tables. Also, some users reported that the navigation in EEJP is not that clear, which might cause negative feelings toward perceived usability of the current system. Moreover, Tsakonas and Papatheodorou (2006) suggested that “Navigation” be one of the five criteria considered when researchers and programmers design and develop an e-learning system or tool, in this regard, the findings here could provide some empirical evidence to clarify the importance of navigation.

The quantitative and qualitative results concerning the factors influencing users’ perceived usefulness, usability, attitude, and follow-up intention of using EEJP were analyzed in the present study. The results show that users’ prior e-learning experience was found to influence users’ perceived usefulness and usability of EEJP, which agrees with those of the empirical studies discussed above (Hurtienne, Horn, Langdon & Clarkson, 2013; Mitzner et al., 2010; Varma & Marler, 2013). These results lend some credence to the hypothesis that users’ prior experience of using e-learning tools tended to impact users’ perceived usefulness and usability (Mitzner et al., 2010). Since EEJP-Write is a relatively complicated tool incorporating many functions, users with more experience and better knowledge and ability would be more satisfied with the learning contents and system functions, which was congruent with the study presenting that nurses with more experience in dealing with electronic patient records had more satisfaction on the adoption to the new management system (Chen & Yao, 2016; Maillet et al., 2015). Therefore, this study recommended learners to accumulate their experience of using e-learning tools, and suggested e-learning tool developers to emphasize more on the instructions regarding the operation toward a complicated e-learning tool before it is actually tried and used.

The present study should lead to a better picture of how GBWI is incorporated into e-learning tools, and could be utilized for EFL graduate students to develop English academic writing proficiency. Implications could also be
applied to improve the teaching pedagogy in the ESP writing composition courses in higher education. In general, such insights would be useful to improve EFL graduate students’ English academic writing ability and particularly to see the usefulness of ESP writing pedagogy. On a more specific basis, the present research could serve to reinforce the usefulness of GBWI for engineering graduate students in Taiwan and to systematize the factors influencing user attitudes and follow-up intention to use e-learning systems and tools. Finally, the present study was conducted with the hope of shedding light on several issues and of paving the way to new research projects which will help consolidate the study of English writing pedagogy in EFL contexts. Future research could also investigate how EEJP might be incorporated into actual educational practice in English academic writing for EFL graduate students. Finally, studies that incorporate EEJP into the blended GBWI courses are recommended to explore the usefulness of GBWI in e-learning tools for NNES engineering graduate students to cultivate academic writing ability in English.

Conclusions

In conclusion, in the present study, an e-learning tool, Engineering English Journal Paper Writing (EEJP), based on the results of a Needs Analysis (NA) and on the Genre-based Writing Instruction (GBWI) approach, was developed and evaluated. A total of ten engineering-based disciplines were selected as the major foci of the design and development of EEJP-write, and 150 selected journals were analyzed as the main learning materials. Moreover, 10 professors wrote sentence examples, after which they were proofread by a professional proofreader who owns specific knowledge of English as a Second Language (ESL), English for Specific Purposes (ESP), and owns content knowledge of engineering-based fields. 32 participants with experience of writing academic works in English volunteered to take part in the trial session, and shared their experiences of using EEJP through filling out the questionnaire and participating in interviews. The results of the present study showed that participants were found to be satisfied with EEJP, and reported that the writing templates and sentence examples were very useful and helpful when writing academic papers in English. Participants also reported needing to spend time becoming familiarized with some functions, which was troublesome since functions such as inserting tables and figures and making formulae were quite different from the methods they were familiar with in the Microsoft platform. Accordingly, they expressed a desire that future system designers could simplify these functions. Through paired-sample t-test, perceived usefulness and usability were found to be influenced by users’ prior experience of using e-learning tools, which was also reported in the findings of previous studies. Concerning the content, users reported the learning materials provided in the proposed system were useful for learning how to write international journal papers in English.

The present findings contribute to the understanding of the various forces acting on the design, development and evaluation of EEJP. One such force is the impact of GBWI on the design and preparation of learning materials for English academic writing with e-learning tools such as EEJP. Through the guidance of the analyzed move structures and sentence examples, users reported that they wrote effectively and efficiently, however, the present study only focused on 20 journal papers in selected indexed journals in each field, which have neglected consideration of subtle differences. Thus, future studies could incorporate more academic papers from different journals in order to have more comprehensive writing templates and research genres.

Acknowledgements

This work was supported by the Ministry of Science & Technology (MOST 102-2511-S-006-005-MY3 & MOST 105-2511-S-006-006-MY3) and the Research Institute for the Humanities and Social Sciences, MOST, Taiwan. This research was also, in part, supported by the Ministry of Education, Taiwan, R.O.C. The Aim for the Top University Project to the National Cheng Kung University.

References


Guest Editorial: Fostering Deep Learning in Problem-Solving Contexts with the Support of Technology

Minhong Wang1*, Sharon Derry2 and Xun Ge3

1KM&EL Lab, Faculty of Education, The University of Hong Kong, Hong Kong // 2School of Education, University of North Carolina at Chapel Hill, NC, USA // 3Department of Educational Psychology, University of Oklahoma, OK, USA // magwang@hku.hk // derry@unc.edu // xge@ou.edu

*Corresponding author

Introduction

Learning and cognition occur in physical and social contexts where knowledge is created and applied, as claimed by situated learning theories (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991). Knowledge is assumed to be better constructed through interaction with problem-oriented, socially situated environments. Accordingly, learning through problem solving in real-world situations, especially with ill-structured problems (Jonassen, 1997) and authentic whole tasks (van Merriënboer & Kirschner, 2013), has become the central aspect of educational practice. The literature has shown the promising effects of problem-oriented learning in helping students to develop critical thinking and problem-solving skills as well as consolidate and extend subject-matter knowledge. Meanwhile, researchers have reported inconclusive and inconsistent findings on the superiority of problem-oriented learning over conventional instructions (Dochy, Segers, van den Bossche, & Gijbels, 2003), mainly in systematic construction of subject-matter knowledge and the development of efficient reasoning process (Coderre, Mandin, Harasym, & Fick, 2003; Patel, Yoskowitz, Arocha, & Shortliffe, 2009).

Given the constraints of classroom settings in enabling situated learning, technology-supported learning environments have been increasingly explored to support learning with real-world problems and authentic tasks in blended environments. Virtual worlds, simulations, and web-based systems have been increasingly employed to expand the opportunities to learn in authentic contexts (Dede, 2009; Derry, Levin, & Schauble, 1995; Linn, 2000). Technology-supported learning environments have shown their advantages in affording flexible access to information and learning resources, on-demand delivery of learning programs, flexible communication and social interaction, effective data processing and other operations, multimedia representations, and more importantly computer-based learning support.

However, effective learning through problem solving is difficult to realize in both classroom and technology-mediated settings. Solving a real-world problem often involves a sophisticated process of understanding the problem, linking abstract knowledge to problem information, and applying relevant methods and strategies to solve the problem. Such a complex process can generate a heavy cognitive load for learners (Kirschner, Sweller, & Clark, 2006), although the complexity of the learning process is often overlooked by instructors or experts, as for them many of the requisite processes have become largely automatic or subconscious with experience. As a result of their limited abilities to deal with complex problem-solving processes, many learners tend to engage in surface rather than deep learning experience that enables them to achieve desired learning outcomes (Wang, Kirschner, & Bridges, 2016).

Deep learning is characterized by a high level of engagement in learning, driven by intrinsic motivation and more importantly, supported by relevant learning approaches or strategies that allow learners to manage complexity and key challenges (most on cognitive aspects) to sustain engagement and achieve a high level of understanding and performance. While deep learning is driven by intrinsic motivation (Biggs, 1993), cognitive approaches are crucial for helping learners persist through challenges and setbacks in to achieve desired learning outcomes. Cognitive approaches to fostering deep learning in problem-solving contexts are associated with multiple issues involving externalizing the tacit aspects of complex tasks for effective thinking, action, and reflection; relating new ideas with prior knowledge and experience for effective construction of knowledge from practice; converging knowledge by resolving conflicts in social contexts; and combining discrete pieces of knowledge into a coherent whole (Chin & Brown, 2000; Entwistle, 2000).

Making the tacit aspects of complex problem-solving tasks explicit and accessible to learners is related to scaffolding, which has been increasingly recognized as important part of learning in problem-solving contexts (Hmelo-Silver, Duncan, & Chinn, 2007). For example, prompts are used to bring learners’ attention to important issues during an ill-structured problem-solving task (Ge & Land, 2003); a complex problem-solving task is decomposed into a set of main actions or key questions to help learners recognize the important goals to pursue during the task (Reiser, 2004). Recent research has highlighted the importance of making thinking visible in
complex problem or task situations (Linn, 2000). Learners’ active construction of external representations related to the solution of a problem has received increased attention. For example, causal maps representing the relationship of cause and effect (Slof, Erkens, Kirschner, Janssen, & Jaspers, 2012), evidence maps linking evidence with claims or hypotheses (Suthers, Vatraru, Medina, Joseph, & Dwyer, 2008), and integrated cognitive maps connecting problem-solving and knowledge-construction processes (Wang, Wu, Kinshuk, Chen, & Spector, 2013) have shown their promising effects in improving understanding and performance in problem-solving contexts.

Making complex thinking visible and accessible is important not only for performing a problem-solving task, but also for constructing knowledge from problem-solving experience, converging knowledge by resolving conflicts in social contexts, and combining discrete pieces of knowledge into a coherent whole. Making thinking visible is much more easily advocated than accomplished (Linn, 2000). More research is needed to examine whether and how deep learning can be fostered from multiple perspectives in problem-oriented, socially situated environments.

Learning through problem solving is not new. It is more important than ever in today’s rapidly changing world, where learners are required to deal with more sophisticated real-world problems, and have more exposure to authentic experience. While technology has substantially expanded the environment for learning with authentic problems, it is critical to bolster the understanding of multiple challenges of learning in problem-solving contexts, and how such challenges can be resolved by effective design and analysis of learning in technology-supported environments. This special issue aims to provide a platform for researchers to present their findings and efforts that may offer insights into how deep learning in problem-solving contexts can be fostered with the support of technology from different perspectives. The focus is on the challenges of learning in problem-solving contexts, effective design of technology-supported learning environments that address the challenges, and meaningful analysis of learning in such environments.

Preview of papers

In the first paper “EcoXPT: Designing for deeper learning through experimentation in an immersive virtual Ecosystem,” Chris Dede, Tina A. Grotzer, Amy Kamarainen and Shari Metcalf present an inquiry-based middle school curriculum that supports blended learning with ecosystem science by utilizing immersive authentic simulation and experimentation together with scaffolding tools in virtual, mixed, and augmented reality environments. Deeper learning in the EcoXPT curriculum is fostered by six strategies, namely case-based instruction, the use of multiple representations of concepts, collaborative learning, apprenticeship-based learning, learning for transfer, and the use of diagnostic assessments.

The second paper “Comparing design constraints to support learning in technology-guided inquiry projects” by Lauren Applebaum, Jonathan Vitale, Libby Gerard and Marcia Linn examined the design and effects of a blended approach to project-based inquiry learning, where hands-on design projects and web-based tools are integrated to improve science learning. Given that many students fail to apply core science principles to their design, web-based models and tools are used to help learners capture such principles during the project. Moreover, a constraint-based goal is imposed on the design project with a view to inducing students’ creativity when they work with constraint-based design problems.

In the third paper “Design of a three-dimensional cognitive mapping approach to support inquiry learning,” Juanjuan Chen, Minhong Wang, Chris Dede and Tina A. Grotzer examined how a novel three-dimensional cognitive mapping (3DCM) approach makes complex inquiry learning visible and accessible to middle school students, by allowing them to externalize the hypothesizing and reasoning process, subject-matter knowledge, and problem information in a single image for effective thinking, action, and reflection. Using this approach, students at a low academic level acquired more knowledge than either the high-level or medium-level students, thus narrowing the academic gap between low-level, medium-level, and high-level students.

Student prior knowledge of scientific phenomenon is often fragmented and plagued with incompatible and incomplete understanding, but valuable for generating curriculum that encourages deeper understanding of scientific concepts. In the fourth paper “Leveraging students’ prior knowledge to adapt science curricula to local context” Lana M. Minshew, Kelly J. Barber-Lester, Sharon J. Derry, and Janice L. Anderson investigated how students’ prior knowledge can be leveraged in curriculum design to promote deeper learning in science curricula. The study presents a model-based assessment that elicits the evidence of student understanding of key concepts and relationships relevant to energy and matter in an ecosystem.
The fifth paper “Moving apart and coming together: Discourse, engagement, and deep learning” by Andrea S. Gomoll, Cindy E. Hmelo-Silver, Erin Tolar, Selma Šabanović, and Matthew Francisco investigated how students collaboratively construct and represent shared understanding in a complex, problem-oriented, and authentic learning environment with a robotics design project. The findings reveal the importance of embodied actions of learning in supporting deep and robust engagement in collaborative learning, for example by positioning authority and accountability, directing attention, and providing support for verbal reasoning.

With limited problem-solving capability and practical experience, it is difficult for novices to develop expert-like performance without necessary support. In the sixth paper “Deep learning towards expertise development in a visualization-based learning environment,” Bei Yuan, Minhong Wang, Andre W. Kushniruk, and Jun Peng examined the design and effects of a model-based learning approach implemented in a web-based learning environment that allows learners to capture and reflect on their problem-solving process in visual formats as well as identify the gap between their performance and that of the expert for effective reflection and improvement towards expertise development.

In the seventh paper “Deep and surface processing of instructor’s feedback in an online course,” Kun Huang, Xun Ge, and Victor Law explored the characteristics of deep and surface approaches to learning online students demonstrated in their responses to instructor’s qualitative feedback to a multi-stage, ill-structured design project. The findings reveal the patterns of deep and surface learning manifested in individual approaches to addressing feedback, and the influence of learner characteristics such as epistemic beliefs and need for closure on individual approaches. The findings provide insights into feedback strategies for deep learning.

Owing to a lack of opportunities for authentic use of a foreign language, many foreign language students fail to use the target language in a meaningful way. The eighth paper “Investigating the effects of authentic activities on foreign language learning: A design-based research approach” by Ildeniz Ozverir, Ulker Vanci Osam, and Jan Herrington presents the design and analysis of a computer-assisted foreign language learning environment, where task-, problem-, and project-based authentic learning activities were incorporated into a Moodle system containing relevant learning resources and functions for communication and assessment.

In the ninth paper “Visualizing the complex process for deep learning with an authentic programming project” Jun Peng, Minhong Wang, and Demetrios Sampson investigated how the complex process of carrying out an authentic programming project can be made visible and accessible to learners. Implementing project-based learning remains a challenge in programming education since advanced programming strategies are implicit and hard to capture, but critical for programming. This study examined the effects of a visualization-based learning environment that externalizes the complex process of applying advanced programming strategies to design and develop artifacts of authentic programming projects.

Problem-based learning often involves a great deal of information searching and selection, where students need to identify useful information to solve a problem. In the tenth paper “Can students identify the relevant information to solve a problem?” Lishan Zhang, Shengquan Yu, Baoping Li, and Jing Wang explored the design of a computer-based assessment system and used the system to assess elementary school students’ ability to select relevant information and search for additional information to solve real-world problems.

The eleventh paper “An eye tracking study of high and low performing students in solving interactive and analytical problems” by Yiling Hu, Bian Wu and Xiaqing Gu presents an eye-tracking study that investigated the use of different information processing strategies by students in solving analytical and interactive problems. The results also examined the differences among high- and low-performing students in their use of information-processing strategies to solve the two types of problems.

In the last paper of the special issue, Kaushal Kumar Bhagata and J. Michael Spector reviewed prior work on using technology as a formative assessment and feedback tool in learning with complex and ill-structured tasks. Their paper “Formative assessment in complex problem-solving domains: The emerging role of assessment technologies” highlights the role of technology-enabled formative assessment in supporting learning in problem-solving contexts. The paper also gives recommendations for further research on using technology to support formative assessment in complex problem-solving domains.

**Conclusion**

We conclude by noting that the papers in this special issue are intended to be representative of ongoing research in fostering deep learning in problem-solving contexts with the support of technology. The international scope
and breadth of the research is distinctive. We hope this special issue will foster further interest in what we believe will become an area of increasing importance.

References


EcoXPT: Designing for Deeper Learning through Experimentation in an Immersive Virtual Ecosystem

Chris Dede*, Tina A. Grotzer, Amy Kamarainen and Shari Metcalf
Harvard Graduate School of Education, Cambridge, MA, USA // chris_dede@harvard.edu // tina_grotzer@harvard.edu // amy_kamarainen@harvard.edu // shari_metcalf@harvard.edu

*Corresponding author

ABSTRACT

Young people now must compete in a global, knowledge-based, innovation-centered economy: they must acquire not just academic knowledge, but also character attributes such as intrinsic motivation, persistence, and flexibility. To accomplish these ambitious goals, the National Research Council (2012) of the United States recommends the use of “deeper learning” classroom strategies. These include case-based learning, multiple representations of knowledge, collaborative learning, apprenticeships, life-wide learning, learning for transfer, interdisciplinary studies, personalized learning, connected learning, and diagnostic assessments. Immersive media (virtual reality, multi-user virtual environments, mixed and augmented realities) have affordances that enhance this type of learning. EcoXPT is an inquiry-based middle school curriculum on ecosystem science that invites students into immersive experimentation with scaffolding tools that support deeper learning. This includes a case-based approach situated in an unfolding eutrophication scenario in which students learn new information from their observations over space and time, speaking with virtual characters in the world, and gathering information in the field guide and other sources. Diagnostic assessments of students’ progress are based on multiple sources, including process data from various types of logfiles. Multiple varied forms of representation convey perceptual, graphical, and experimental data, enabling students to investigate relationships between variables. Students are apprenticed in the ways of knowing of ecosystems scientists, which involves interdisciplinary knowledge. Students collaborate in teams of two, subdividing the tasks of gathering evidence.

Keywords

Immersive learning, Virtual worlds, Deeper learning, Ecosystems science

Deeper learning to prepare young people for life and work in the 21st century

Young people now must compete in a global, knowledge-based, innovation-centered economy (Araya & Peters, 2010). In order to secure a reasonably comfortable lifestyle, they now must go beyond a high school diploma (Wagner, 2008), and they must acquire not just academic knowledge, but also character attributes such as intrinsic motivation, persistence, and flexibility (Dede, 2010; Levin, 2012; National Research Council, 2008). As described by the National Research Council (NRC) in its landmark report Education for Life and Work in the 21st Century (2012), cognitive, intrapersonal, and interpersonal dimensions of knowledge and skills are best developed in combination. Table 1 categorizes a broad range of knowledge and skills vital in the 21st century, according to these dimensions. Moreover, and in contrast to industrial-era schooling with its emphasis on multiple choice and short-answer testing, mastery now requires the ability to apply knowledge and skills in real-world contexts, demonstrating proficiency via effective, authentic performances (Dede, 2014).

<table>
<thead>
<tr>
<th>Cognitive outcomes</th>
<th>Intra-personal outcomes</th>
<th>Inter-personal outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive processes and strategies</td>
<td>Intellectual openness</td>
<td>Teamwork</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Work ethic and conscientiousness</td>
<td>Leadership</td>
</tr>
<tr>
<td>Creativity</td>
<td>Positive core self-evaluation</td>
<td>Communication</td>
</tr>
<tr>
<td>Critical thinking</td>
<td>Metacognition</td>
<td>Responsibility</td>
</tr>
<tr>
<td>Information literacy</td>
<td>Flexibility</td>
<td>Conflict resolution</td>
</tr>
<tr>
<td>Reasoning</td>
<td>Initiative</td>
<td></td>
</tr>
<tr>
<td>Innovation</td>
<td>Appreciation of diversity</td>
<td></td>
</tr>
</tbody>
</table>

For all students to reach such ambitious standards, not just an elite few, how must schools change? In order to make these types of learning outcomes possible, on a large scale, what kinds of instruction would have to become common practice?

To accomplish these ambitious goals, the 2012 NRC report recommends the use of “deeper learning” classroom strategies. The approaches promoted by advocates of deeper learning are not new, and historically these instructional strategies have been described under a variety of terms. Until now, however, they have been rarely practiced within the industrial era schools.

ISSN 1436-4522 (online) and 1176-3647 (print). This article of the Journal of Educational Technology & Society is available under Creative Commons CC-BY-ND-NC 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Journal Editors at ets-editors@ifets.info.
- *Case-based learning* helps students master abstract principles and skills through the analysis of real-world situations;
- *Multiple, varied representations* of concepts provide different ways of explaining complicated things, showing how those depictions are alternative forms of the same underlying ideas;
- *Collaborative learning* enables a team to combine its knowledge and skills in making sense of a complex phenomenon;
- *Apprenticeships* involve working with a mentor who has a specific real-world role and, over time, enables mastery of their knowledge and skills;
- *Self-directed, life-wide, open-ended learning* is based on students’ passions and connected to students’ identities in ways that foster academic engagement, self-efficacy, and tenacity;
- *Learning for transfer* emphasizes that the measure of mastery is application in life rather than simply in the classroom;
- *Interdisciplinary studies* help students see how differing fields can complement each other, offering a richer perspective on the world than any single discipline can provide;
- *Personalized learning* ensures that students receive instruction and supports that are tailored to their needs and responsive to their interests (U. S. Department of Education, 2010; Wolf, 2010);
- *Connected learning* encourages students to confront challenges and pursue opportunities that exist outside of their classrooms and campuses (Ito et al., 2013); and
- *Diagnostic assessments* are embedded into learning and are formative for further learning and instruction.

These types of learning entail very different teaching strategies than the familiar, lecture-based forms of instruction characteristic of conventional schooling, with its one-size-fits-all processing of students. Rather than requiring rote memorization and individual mastery of prescribed material, these approaches involve in-depth, differentiated content; authentic diagnostic assessment embedded in instruction; active forms of learning, often collaborative; and learning about academic subjects linked to personal passions and infused throughout life (Dede, 2014).

### Designing immersive authentic simulations to promote deeper learning

As can be seen from the list above, deeper learning experiences designed to teach complex knowledge and sophisticated skills are often based on “guided social constructivist” theories of learning. In this approach, learning involves mastering authentic tasks in personally relevant, realistic situations. Meaning is imposed by the individual rather than existing in the world independently, so people construct new knowledge and understandings based on what they already know and believe, which is shaped by their developmental level, their prior experiences, and their sociocultural background and context (Palincsar, 1998). Instruction can foster learning by providing rich, loosely structured experiences and guidance (such as apprenticeships, coaching, and mentoring) that encourage meaning-making without imposing a fixed set of knowledge and skills. This type of learning is usually social; students build personal interpretations of reality based on experiences and interactions with others.

Immersive media have affordances that enhance this type of learning. Psychological immersion is the mental state of being completely absorbed or engaged with something. For example, a well-designed game in a multi-user virtual environment (MUVE) draws viewers into the world portrayed on the screen, and they feel caught up in that digital context. The use of narrative and symbolism creates credible, engaging situations (Dawley & Dede, 2013); each participant can influence what happens through their actions and can interact with others. Via richer stimuli, head-mounted or room-sized displays can create sensory immersion to deepen the effect of psychological immersion, as well as induce virtual presence (place illusion), the feeling that you are at a location in the virtual world rather than the place where your physical body is (Slater, 2009).

### EcoMUVE as an immersive authentic simulation

Immersive virtual learning environments can enhance learning of science concepts by situating students’ investigations in realistic, yet scaffolded contexts (Colella, 2000; Dawley & Dede, 2013; Ketelhut et al., 2010). Situated experimental tools let students interpret results contextually and integrate their findings with other sources of evidence--including observations and data collected in the virtual world--to build and test hypotheses.

EcoXPT is based on a prior immersive authentic simulation project, EcoMUVE. The EcoMUVE middle grades curriculum focuses on the potential of immersive authentic simulations for teaching ecosystems science
EcoMUVE is an ill-structured problem space in which students collaborate to construct meaning about scientific phenomena embedded in the immersive pond or forest environment (Kamarainen, Metcalf, Grotzer, & Dede, 2015). Within the immersive virtual environments, students have multiple ways of accessing information about the relationships among components of the system, students find opportunities to collect multiple forms of evidence, and the rich relationships among sources of evidence render these open to various interpretations. Thus, the complexity of the immersive world provides a context in which students must justify their interpretation of the relationships as they build and revise a conceptual model of these relationships during an ongoing concept mapping activity. Using immersion to position students within an ill-structured problem space helps them engage in collaborative sense making that encouraged use of evidence in support of claims. The behaviors of students included collecting data, distinguishing between problem-relevant and problem-irrelevant information, and collaboratively reasoning about the claims represented in their concept maps by combining prior knowledge and repeated visits to the virtual environment to revise their understanding. Thus, immersion in the virtual world, supporting design and curricular features, and the concept mapping task elicited behaviors that are closely aligned with the epistemic work of experts in the field of ecosystem science.

Prior research with EcoMUVE demonstrated significant student gains in ecosystem science knowledge (Kamarainen et al., 2012; Metcalf et al., 2011) and in causal understanding (Grotzer et al., 2013). A study on teacher perceptions of EcoMUVE in the classroom surveyed 16 teachers who had used the curriculum with their students, about the value, effectiveness, and feasibility of EcoMUVE based on their experiences; some teachers additionally participated in a comparison study of EcoMUVE with a non-MUVE curriculum (Kamarainen et al., 2012). Teachers felt EcoMUVE was effective, aligned well with standards, and compared favorably with a non-MUVE alternative. Particular technological and curriculum features that were identified by teachers as valuable included both technological aspects, such as immersion in the virtual environment and easy-to-use data collection and analysis tools, and also pedagogical features, particularly the opportunity for self-directed learning by students and the inquiry, role-based pedagogy (Metcalf et al., 2013).

In a study looking at changes in student motivation as a result of using EcoMUVE (Chen, Metcalf, & Tutwiler, 2014), quantitative data indicated that students’ interest in science did not change from pre-intervention to post-intervention. However, a closer analysis revealed that students who identified more strongly with science did become more interested in science, whereas those who did not identify with science evinced no change in interest for science. A companion study (Metcalf et al., 2014) showed that, over the two-week EcoMUVE curriculum experience, student interest in EcoMUVE decreased somewhat but remained high; students’ beliefs about EcoMUVE’s utility increased; and students saw EcoMUVE as less of a “waste of time.” Student responses to questions of what they liked about EcoMUVE changed from being primarily about the opportunity to interact in a virtual computer environment to an increasing appreciation of the pedagogical aspects of the self-directed, collaborative, inquiry-based activities. These findings demonstrate that, although there is a novelty effect for EcoMUVE, engagement didn’t ultimately depend on novelty. This is an important contribution to the teaching and learning of science because it demonstrates and reinforces the importance of sound pedagogical methods. As technology becomes more prevalent in science classrooms, this study serves as a reminder that, regardless of the medium, it is fundamental to design the technology to allow for active, collaborative student involvement in inquiring about scientific phenomena.

EcoMUVE is a first-generation ecosystems science curriculum that focuses on the observational methods of ecosystems science. We designed and developed EcoXPT (described below) as a second-generation, more complex curriculum that builds on EcoMUVE, but adds new components to the digital ecosystem and includes six types of experimental tools authentic to ecosystems science. The initial pilot of EcoXPT in a few classrooms has just been completed, and data has not yet been analyzed; large-scale trials of EcoXPT will take place in the 2017-18 school year. For this special issue focusing on design, the emphasis is on showing how EcoMUVE and EcoXPT incorporate design for deeper learning, rather than on efficacy studies of the extent to which this design succeeds.
EcoXPT extends EcoMUVE with authentic experimentation in ecosystems science

As a follow-on to EcoMUVE, EcoXPT is an inquiry-based middle school curriculum (see http://ecolearn.gse.harvard.edu/ecoXPT/overview.php) that extends the EcoMUVE Pond curriculum by adding experimental methods authentic to ecosystem science to complement observation and correlation, enabling students to reason about causes. Students investigate why all of the large fish in a virtual pond have died (Figure 1). In addition to the environmental features in EcoMUVE Pond, a farm, a second pond, a second housing development, and a drainage ditch have been added. While immersed in the virtual world, students can assess the slope of the land, thereby gaining a tacit understanding of landscape topography and the boundaries between watersheds. They can further explore the influence of slope on the transport of materials through the ecosystem by using a “tracer tool” to track where water, and the fertilizer it carries, flows on a rainy day. Water flows in a variety of directions, and determining this becomes important in understanding the causes of the fishkill.

Students collect environmental and population data over time, and conduct a variety of experiments in the virtual ecosystem. Teams of students construct hypotheses, supporting their arguments with data and experimental results.

Reasoning in purely observational environments, like EcoMUVE Pond, tends to be inferential, based on visual information, measurements and correlations observed over time. In EcoXPT, we are studying whether the ability to conduct experiments in order to test various causal hypotheses about variables and relationships will support students in deeper, evidence-based reasoning. Further, by situating the simulated experimental tools within the context of the virtual ecosystem, we are studying whether students can adopt more sophisticated approaches to investigation that mirror the thinking moves that ecosystem scientists use when investigating environmental problems. We hypothesize that EcoXPT will give students opportunities to extend their observational learning by applying their experimental findings to building hypotheses about the ecosystem scenario.

EcoXPT invites students into immersive experimentation with multi-dimensional experiences that support deeper learning. Consistent with the instructional strategies of deeper learning, this includes a case-based approach situated in an unfolding eutrophication scenario in which students learn new information through their observations over space and time, from virtual characters in the world, and from gathering information in the field guide and other sources. Multiple varied forms of representation convey perceptual (Figure 2), graphical (Figure 3), and experimental data (Figure 4) to the students, enabling them to investigate relationships between variables. The use and interpretation of varied forms of representation is further supported by the EcoXPT
notebook and concept mapping tools, which help students integrate various forms of evidence into their scientific arguments.

Figure 2. The Submarine tool can be used to observe and measure populations of microscopic organisms.

Figure 3. The Data Visualization tool supports viewing tables and graphs of up to five variables.
EcoXPT illustrates the capacity of immersive simulations to support a wide range of instructional strategies for deeper learning, through design that enables extended, multi-dimensional, authentic experiences.

**EcoXPT enables students to use the tools and the inquiry practices of ecosystem scientists**

Part of deeper learning is helping students to understand that science is not just a body of facts to be memorized, but also a set of practices used collaboratively to expand on what is known about the natural world. EcoXPT supports this integration of knowledge and practice by promoting engagement in use of scientific tools, while aiding reflection on the value of using those tools in conjunction with larger investigative strategies in order to develop and support an evidence-based argument.

EcoXPT tools that illustrate situated experimentation include:

- **Tolerance Tanks** – Students adjust the levels of an environmental factor, such as fertilizer, in three tanks containing different species of fish (e.g., minnow, bluegill), to determine the level (if any) at which each type of fish is affected by that factor (Figure 5).
- **Tracers** – Chemical markers show the movement of matter in the environment. Adding tracers to bags of fertilizer lets students test how the spatial layout and topography affect fertilizer runoff when it rains (Figure 6).
- **Comparison Tanks** - Students configure one to three tanks with experimental factors, and collect measurements – e.g., discovering that a tank with algae has a higher dissolved oxygen than one without algae (Figure 7).
- The **Weather Simulator** enables students to model the effect of various environmental factors (e.g., wind, temperature) on the level of dissolved oxygen in water (Figure 8).
- **Mesocosms** are containers in an outside location that provide a similar, but smaller, environment to the pond (Figure 9). This enables the investigator to assess the impact of variables of interest (e.g., addition of fertilizer) in addition to environmental factors (e.g., sunlight, weather) in ways that more closely mirror the pond than small, indoor tolerance and comparison tanks can.
- The **Sensor Buoy** immersed in the pond provides a steady stream of data about its dynamics over time. (Figure 10).

Figure 4. The Notebook tool stores observations, graphs, testimony, and experimental data, with student annotations.
Figure 5. Tolerance Tank tool

Figure 6. Tracer tool
Figure 7. Comparison tank tool

Figure 8. Weather simulator
These six tools provide a progression that initially enables students to test their hypotheses about simple relationships, then gradually introduces more sophisticated, multidimensional measurements that reveal the complex causality in the pond’s dynamics over time.

In order to learn the thinking inherent in moving from seeking patterns to analysing causality, Thinking Moves accompany these experimental tools in order to help students understand the kinds of questions that ecosystems scientists might ask as they explore the potential causal dynamics of an ecosystem (Grotzer, Kamarainen, Metcalf, Tutwiler, & Dede, 2017):

- **Deep Seeing** encourages students to consider the natural history of the ecosystem and to engage in careful observation of what is there. It asks them to look while being careful to set their assumptions aside.
- **Evidence Seeking** encourages students to collect evidence from multiple sources, to seek corroborating evidence, and to evaluate the sources of their evidence.
- **Pattern Seeking** encourages students to notice patterns in the on-going processes and steady states of the system and to notice how certain variables change together or not.
- **Analyzing Causality** asks students to use experimental evidence and intervention to try to impact change in the patterns in an effort to discern the underlying mechanisms at work.
• **Constructing Explanations** encourages students to develop the best “story” or explanation that they can from the available evidence. It asks students to look for gaps in their explanation and to assess their explanations against rival explanations.

Studies on the effectiveness of these supports are now being conducted.

**How the design of EcoXPT fosters deeper learning**

Consistent with instructional strategies of deeper learning, EcoXPT presents an unfolding eutrophication scenario in which students learn new information from their observations over space and time, speaking with virtual characters in the world, and gathering information in the field guide and other sources. The complexity of the situation grows over time and is shaped by the decisions students make about what data to collect and what experiments to do. This pedagogical method is consistent with case-based teaching and problem-based learning, which extensive research has shown leads to outcomes characteristic of deeper learning (Lu, Bridges, & Hmelo-Silver, 2015).

The experimental tools in EcoXPT provide a progression authentic to ecosystems science that aids the students in understanding the unfolding and increasing complexity of the eutrophication process. Initially, the tolerance tanks and the tracers provide evidence about the role of individual variables and environmental dynamics. Then the comparison tanks and the weather simulator show the interactions among variables. The mesocosms indicate the interactions between suites of variables and environmental dynamics. Finally, the buoy provides a detailed stream of longitudinal diurnal data. This suite of tools can be leveraged by students throughout their investigation to support their understanding of ecosystem science epistemology and experimental processes, as well as building their inquiry skills and their knowledge of eutrophication.

In terms of design for deeper learning, students collaborate in teams of two in a simulated setting designed for transfer of knowledge and skills to real world activities. Intra-team and whole class collaboration aids in both learning and engagement, as students socially construct knowledge and often provide complementary contributions in comprehending various types of representations. Scientists and a teenage virtual guide in the immersive world apprentice students in the ways of knowing of ecosystems scientists, which involves interdisciplinary knowledge. Overall, this is like an internship experience in ecosystems science, focusing on authentic epistemological methods that go beyond simply controlling all but one variable in a situation. To aid students in understanding the complex causality involved, diagnostic assessments of students’ progress are based on teachers’ observations of students’ activities and discussions, in-world and in-classroom, as well as the notebooks and concept maps as ways of making the progression of students’ thinking visible.

**Early research results on effectiveness**

A full version of EcoXPT has just been completed, and studies of its usability and effectiveness in classroom settings are now being conducted. Below are summarized findings related to deeper learning from pilot studies on earlier, partial versions of the curriculum.

In 2016, a pilot study of EcoXPT was conducted with four teachers who taught a total of 14 classes of 7th grade students (N = 280) using an early version of the 2.5 week curriculum (Metcalf et al., 2016; Metcalf et al., 2017; Thompson et al., 2016a). Data collected included a pre-post content survey measuring content knowledge in ecosystems science, consistent with the US Next Generation Science Standards for this grade level, as well as the students’ final presentations, which included concept maps hypothesizing causal relationships. This pre-post content survey (Thompson et al., 2016a) found statistically significant gains (t(288) = 9.5045, p < .001) in ecosystem science content knowledge. This provides evidence that the deeper learning processes used (described in the section above) were successful in enhancing students’ knowledge of ecosystems science concepts, as well as their understanding of complex causality.

Quotes from student presentations demonstrate ways in which students drew conclusions from their experiments.

- “An experiment we did in the mesocosms showed that the bacteria ate the dead matter which is the dead plants. Bacteria is able to do respiration so the bacteria uses the dissolved oxygen. So the bacteria use all of the limited dissolved oxygen, so both the bacteria and the fish die.”
- “We used tracers to trace the fertilizer, the fertilizer ran off of the ground and into the pond. When the fertilizer got to the pond it caused the algae to grow, because the algae are plants.”

Quotes from student presentations demonstrate ways in which students drew conclusions from their experiments.
Concept maps representing teams’ hypotheses illustrate sophisticated learning about the ecosystem relationships (Figure 11). Further, student presentation slides show both situated learning and integrated reasoning with data. This type of evidence-based reasoning indicates the success of our deeper learning processes in achieving strong student outcomes.

![Figure 11. The digital tool for building concept maps](image)

Two studies using mixed-methods analyses (including log file data, video, and think-aloud interviews) about students’ use of and strategies toward experimentation using a set of virtual experimental fish tanks revealed diverse strategies for experimentation (Thompson et al., 2016b) and found that students used more expert strategies for experimentation over time (Metcalf et al., 2016). These cases provided a rich basis for developing supports for effective design of virtual experiments that support development of both conceptual knowledge related to important ecosystem relationships, as well as support for more effective and appropriate strategies for experimentation. Again, this is indicative of the success of our deeper learning processes designed into EcoXPT.

**Conclusions**

Overall, our research suggests that immersive experimentation with scaffolding tools integrated with multiple evidence sources can foster deeper learning about ecosystems. This may be because immersive authentic simulations like EcoXPT offer powerful support for six classroom practices known to lead to deeper learning outcomes: case-based instruction, the use of multiple representations, collaborative learning, apprenticeship-based learning, learning for transfer, and the use of diagnostic assessments.

**Acknowledgements**

This work is supported by the National Science Foundation Grant No. 1416781 to Tina Grotzer and Chris Dede. All opinions, findings, conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation. For more information, visit: http://ecolearn.gse.harvard.edu
References


Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), 3549–3557.


Comparing Design Constraints to Support Learning in Technology-guided Inquiry Projects

Lauren R. Applebaum, Jonathan M. Vitale, Elizabeth Gerard and Marcia C. Linn*
University of California, Berkeley, CA, USA // lrapplebaum@gmail.com // jonvitale@gmail.com // libbygerard@berkeley.edu // mclinn@berkeley.edu

*Corresponding author

ABSTRACT
Physical design projects are a way to motivate and engage students in authentic science and engineering practices. Web-based tools can support design projects to ensure that students address and reflect upon critical science concepts during the course of the project. In addition, by specifying challenging design goals that require students to consider potential trade-offs between features, web-based tools may promote more deliberative scientific inquiry than open-ended or feature maximization goals. To study the role of web-supported projects, we developed an online curriculum that guides students through the planning, building, and analysis of self-propelled vehicles. To address content related to energy transformation we incorporated virtual models that display dynamic graphs of energy levels as a virtual scooter travels along a path. We compared two design goals for the project with different constraints. In the target version students are prompted to build virtual and physical scooters to reach a specific position. In the distance version students are prompted to maximize the distance the scooter travels. Our results indicate that students learned energy concepts from both versions; however, students with the target version did refer to the virtual model in their posttest responses to a greater degree than those with the distance version.

Keywords
Blended learning, Technology-enhanced instruction, Project-based learning, Science education

Introduction
In this study we investigate a blended, physical–virtual approach to inquiry learning projects that takes advantage of the motivational affordances of hands-on projects, as well as the guidance and scaffolding affordances of online learning. In particular, new educational standards (e.g., NGSS Lead States, 2013) recognize that hands-on projects can support mastery over disciplinary practices and understanding of core disciplinary ideas in an engaging context. The process of building a working model or artifact requires the learner to generate, integrate, and apply scientific ideas to solve real-world problems. Yet, despite this potential, open-ended design projects often draw students’ attention to superficial structural or aesthetic issues, rather than underlying behaviors and functions (Hmelo, Holton, & Kolodner, 2000). On the other hand, online learning tools can structure content to make underlying mechanisms concrete (Edelson, Gordin, & Pea, 1999; Reiser, 2004). Furthermore, with adaptive mechanisms, online learning tools can provide guidance tailored to students’ individual strengths and challenges (Linn et al., 2014). Bringing together physical and virtual modalities of inquiry, therefore, represents an opportunity to deliver on the enormous, but mostly untapped potential of hands-on projects in science classrooms.

Determining how to best coordinate between hands-on design and virtual tools is an open research question. We explore how a blended inquiry project may be designed so that virtual and physical components build upon each other and help students explore the underlying scientific mechanisms. In this paper, we combine a virtual and physical model in an inquiry activity to explore trade-offs between features of the design. Specifically, students build a self-propelled vehicle and use it to explore issues of energy conservation and transformation. In the target condition, students refine their design so that the vehicle reaches, but does not go beyond, a target. In the distance condition, students refine their design so that their vehicle goes as far as possible. The distance condition is consistent with students’ typical goals for a vehicle. We discuss how both approaches take advantage of the virtual model and physical design to support learning. We hypothesize that the target condition will be better at helping students deliberately distinguish among their ideas.

Challenges with projects
Most students enjoy projects, but often fail to learn and apply core science principles to improve their designs (Crismond, 2011; Hmelo et al., 2000; Horn, 1922; Kanter, 2009; Karaçalli & Korur, 2014; Larmer & Mergendoller, 2010). Rather, in many cases, projects are implemented as arts and crafts activities devoid of science content (Larmer & Mergendoller, 2010). Many teachers avoid implementing hands-on projects, in favor of typical, lecture-based instruction.
Thus, engineering projects where students explore a distance goal may lead to different patterns of exploration than a project where students are encouraged to explore each variable in service of reaching a goal. For example, in the topic explored here, students are prompted to build a self-propelled vehicle using either a balloon or a wind-up device. A typical criterion for the success of their construction is the distance that the vehicle travels before stopping. We compare a typical distance condition to a target condition where students aim to get the self-propelled vehicle to stop at a target.

Supporting knowledge integration with web-based tools

While it is difficult for a single teacher to guide all students during project-based inquiry (Özel, 2013; Tal, Krajcik, & Blumenfeld, 2006), web-based tools can make projects more successful by augmenting teacher guidance. Research on adaptive guidance with digital inquiry tools (e.g., Gobert, Sao Pedro, Raziuddin, & Baker, 2013; Leelawong & Biswas, 2008; Linn et al., 2014; Roscoe & McNamara, 2013), suggests that adaptive technologies can assist students in learning scientific concepts through inquiry. Although well-defined environments, with clear expectations for learning, are the most straightforward approaches to develop adaptive guidance, new approaches for guidance and formative assessment are emerging. Designing activities to take advantage of adaptive capabilities of online systems is an important challenge for educational research.

The knowledge integration (KI) framework (Linn & Eylon, 2011) grew out of research showing that learners have multiple, often conflicting, ideas about scientific phenomena and that they benefit from building on their ideas. When new ideas are added and not distinguished from prior ideas, students often use the new idea in the situation but revert to their earlier ideas subsequently. Thus, to promote KI, guided inquiry is advantageous. The KI framework emphasizes four reasoning processes. First, instruction needs to elicit student ideas so they can be considered when new ideas are added. Second, adding new ideas using interactive virtual models can be valuable. Third, learning occurs when students distinguish new ideas from their prior knowledge and use evidence from their explorations to form a new understanding. Accordingly, in addition to exploratory activities in which students may encounter new ideas, knowledge integration activities encourage exploration of students’ elicited ideas. Fourth, students benefit from reflecting on the process of comparing ideas and consolidating understanding. Guidance can help students distinguish and compare their ideas to form a coherent understanding. While knowledge integration can be supported in all forms of instruction, the Web-based Inquiry Science Environment (WISE) was designed to use logs of student activities to offer adaptive guidance and to report student ideas back to their teachers (Linn et al., 2014).

The KI approach focuses instructional design on distinguishing ideas. It calls for implementing a narrative or goal that binds activities and drives inquiry (Linn & Eylon, 2011). Furthermore, this narrative or goal must reflect inherent difficulties in the material. For example, in a KI activity addressing the question of “what makes an object float?” students are guided to generate virtual artifacts of varying masses and volumes to test their hypotheses (Vitale, Madhok, & Linn, 2016). Because students’ initial ideas are often incomplete (e.g., only mass determines flotation), exploration provides surprising results, and in turn motivates further investigation of underlying mechanisms. However, developing a self-propelled vehicle to maximize distance travelled may not invoke a challenge to students’ ideas. Simply through experience with related materials (e.g., toys, carts), students likely know that minimizing friction between the axle and the body of the scooter (i.e., enabling the wheel to spin freely) and maximizing initial energy will result in the best scooter. Therefore, another approach to structuring the engineering task may be required to encourage exploration of more complex relationships.

Design constraints and knowledge integration (KI)

To facilitate students distinguishing ideas, we compare two design constraints in this study. The target condition has the constraint of hitting a target in both physical and virtual model-testing. Students must design a scooter that lands on the target to achieve the goal. To do this, they can manipulate a number of factors, including mass, friction, and wheel size. The introduction of this design constraint, i.e., a target, may emphasize the trade-offs between mass and friction. In the distance condition, the design constraint is to make the scooter go as far as possible. This design might lead to fewer explorations of trade-offs. Focus on design trade-offs could support more systematic exploration of the variables by students. The target condition might motivate systematic investigation of underlying science issues to seek an acceptable balance between variables (Barron et al., 1998; Kanter, 2009). For example, imposing specific specifications on architectural design (i.e., specifying limitations), requires creative use of tools and materials to produce a structure (Kuhn, 2001).
In terms of knowledge integration, the target constraint could encourage learners to distinguish between the effects of the variables more than the distance constraint. We compare these constraints in the study reported here.

This study

A central science concept for middle school students is the conservation of energy – i.e., that energy in a system is never lost but transforms from one form to another (NGSS Lead States, 2013). However, this concept is challenging, particularly when energy transformation occurs between forms that are easily detectable (e.g., movement, temperature) and tacit forms (e.g., potential energy) (Edens & Potter, 2003; Linn, Songer, Lewis, & Stern, 1993; Liu & McKeough, 2005). Thus, understanding that the movement of a vehicle from start to stop is the result of transformation from potential to kinetic to thermal energy (from friction), with a constant amount of total energy, is not likely to be intuitive for most students. In this potential-kinetic-thermal (“PKT”) system, losses in one form of energy result in equivalent gains across some distribution of the other two energy forms.

While functional physical models (i.e., self-propelled vehicles) inherently demonstrate mechanisms of energy transformation and conservation, these mechanisms may not be clear to students. In particular, prior to (and often after) instruction, many children believe that moving objects naturally come to a stop without any external force or source of energy transfer (McCluskey, 1983). Virtual models, depicting the transformation of energy implicit within the motion of a cart, can help illustrate these concepts and prepare students to understand their physical counterparts.

Yet, simply illustrating relationships does not take advantage of the affordances of design tasks and games to align task goals and learning objectives. While a student could analyze changes in energy after successfully achieving the goal of maximizing the distance the car travels, nothing compels him or her to do so. On the other hand, a goal that imposes constraints – i.e., make the car go a specific distance – may require analysis of energy transformation in order to complete the task successfully.

Accordingly, in the current study, students engaged in a 10-day design project where they would explore virtual and physical models of self-propelled vehicles with either a distance goal (i.e., maximize distance) or a target goal (i.e., hit a target at a specified distance). The aim of the current study is to assess the impact of these two conditions on student performance and on learning of science concepts. Therefore, we ask the following questions:

- Does the Self-propelled Vehicles curriculum unit support learning of energy concepts?
- How do students’ exploration strategies differ by condition?
- Do learning gains in Self-propelled Vehicles differ by condition?
- How do students’ experiences exploring and building models in each condition impact the evidence that they use to justify their responses at posttest?

Method

Participants

This study was conducted with two teachers and their 228 eighth grade (13-14 year-old) students (A: male, 10+ years teaching experience, 103 students; B: female, 2nd year teacher, 125 students) from a single school. This sample accounted for nearly the entirety of the 8th grade class (except for one additional class, taught by a third teacher, whose data is not used in this study). This school serves a diverse, but mostly mid-to-high socioeconomic status, suburban population in the western United States (38% White, 31% Asian, 17% Hispanic, 4% Black, 22% Reduced-price lunch, 12% ELL).

Materials

Assessments

A pretest consists of three sets of multi-part items addressing the relationship among different types of energies (Energy Conservation Graphs, Scooter Revision Graphs, Car Performance Comparison). A posttest consists of these three items, along with an additional item (Consultant). For each of these multi-part items, we chose to
evaluate a single response component that best highlighted students' thinking about energy. The posttest may be viewed at this link: http://wise.berkeley.edu/project/18843#/vle/node1

**Energy Conservation Graphs (henceforth “Conservation”).** For this item, students are presented with four bar graphs that depict possible relations between potential, kinetic, and thermal energy while a scooter is moving. Only one graph correctly displays a total of 100% energy distributed among available energy types (Graph 1d). Students are prompted to choose the correct graph and explain their selection. We coded and evaluated responses to this final prompt.

![Figure 1. Response options for Energy Conservation item](image)

**Scooter Revision Graphs (henceforth “Revision”).** In this item students are presented with a stacked bar graph depicting the distribution of energy (like Figure 1) from an arbitrary timepoint during a scooter’s run. This item also presents two fictional students’ plan for revising their scooter to increase kinetic energy. “Jaden” plans to reduce axle friction, while “Jordan” plans to make the wheels bigger. Students are prompted to choose whether they agree with Jaden (correct), Jordan, both, or neither. After making their selection, students are prompted to reflect upon the stacked bar graph to indicate which form(s) of energy will NOT change based upon this scooter revision. In this case, the potential energy will not change because adjusting friction only impacts the relative distribution of energy between kinetic and thermal. Students are also prompted to explain their reasoning, which we coded and evaluated.

**Car Performance Comparison (henceforth “Comparison”).** In this item, students are prompted with, “Liz and Destiny each built a rubber band car. Liz’s car had more potential energy, but Destiny’s car travelled a farther distance.” Students are instructed to present reasons why Liz’s car did not go as far, even though it had greater potential energy. Following this prompt, students are instructed to choose one of these reasons and explain how the selected reason affected how Liz’s car transformed potential energy to kinetic and thermal energy. We coded and evaluated responses to this final prompt.

**Consultant (posttest only).** This item presents an image of “Lorena’s” scooter and text indicating that the scooter stopped in the middle of the track. Students are prompted to list changes that can be made to the scooter to increase the distance it travels. We coded and evaluated responses to this listing prompt. Following this, students are prompted to choose one of their proposed changes and explain how the change would affect the scooter’s energy.

**Curriculum materials**

All curricular materials are presented in the WISE unit, *Self-propelled Vehicles* (henceforth *Scooters*). During *Scooters* students are introduced to energy concepts of transformation and conservation, and then have the opportunity to apply these concepts to an engineering task. For a layout of the *Scooters* unit, see Table 1. The target version of the curriculum may be found at the following web address: http://wise.berkeley.edu/project/18684#/vle/node1
Table 1. Layout of Scooters lesson

<table>
<thead>
<tr>
<th>Activity number</th>
<th>Activity Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction to Scooters</td>
</tr>
<tr>
<td>2</td>
<td>Design Process 1</td>
</tr>
<tr>
<td>3</td>
<td>Energy Concepts (with virtual models)</td>
</tr>
<tr>
<td>4</td>
<td>Design Process 2</td>
</tr>
<tr>
<td>5</td>
<td>Reflection</td>
</tr>
</tbody>
</table>

Introduction to Scooters. Students begin the Scooters lesson with an introduction to what self-propelled vehicles look like and then they are prompted to describe what features might be important for a scooter to have in order for it to move. This activity does not differ by condition.

Design process 1. Following the introduction, students are presented with instructions on how to design a base scooter model. By starting the students with predefined scooter-building instructions, students can immediately have a successful experience building and testing a functioning scooter. This activity does not differ by condition.

Energy concepts (with virtual models). In this activity students are introduced to relevant energy concepts of transformation and conservation. Central to this activity, students explore a virtual model of a balloon-powered, self-propelled vehicle, whose features can be manipulated (Figures 2 and 3). Specifically, students can manipulate four of the scooter’s features: amount of air in the balloon, mass, friction, and wheel size. In this particular model, all features except for the wheel size affect how far the scooter will travel. Taking advantage of the virtual environment, the results of each model trial are depicted as graphs. On the left-hand side, the distance vs. time information is displayed in real time as the scooter moves across the screen. On the right-hand side, energy distribution is depicted as a stacked bar graph summing to 100% at all times. This graph changes dynamically as the scooter travels.

Depending on experimental condition, students are presented with differing instructions and alternative versions of the virtual model. In the distance condition (Figure 2), students are informed that their goal when revising their physical scooters will be to “have your scooter travel the farthest distance.” Students are encouraged to explore the model in preparation for the goal.

On the other hand, in the target condition (Figure 3), students are informed that their goal when revising their physical scooters will be to reach a target that might be a farther or shorter distance than their initial scooter travelled. In addition, in the virtual model students are presented with a “target” feature, and asked to adjust parameters so that the scooter reaches the target. In the model, once a target is successfully landed upon (within a
margin of error equivalent to half the length of the scooter), the target is automatically moved to a new, random location along the track.

During the course of this activity, students interact with three different virtual versions of either the target or distance model. In all versions, the model input parameters remain constant; however, the right-hand stacked-bar energy graph is updated to reflect the introduction of new concepts. In particular, in the first model (Figure 4) only the potential energy is displayed and labelled, while all remaining energy (summing kinetic and thermal) is labelled as “other”. In the second model, once the concept of kinetic energy is introduced, the energy graph distinguishes between potential and kinetic energy (blue and green, respectively), while the remaining energy is labelled as “other”. Finally, in the last model all three forms of energy are displayed and labelled. By limiting the forms of energy displayed in the first two models, we aim to focus students’ attention on a single form of energy at a time, while also motivating the introduction of new forms of energy with the “other” bar.

**Design Process 2.** After exploring the computer models, students in both conditions are guided to re-design, re-build, and re-test their original scooters. Unlike the first design, the goal of this engineering process differed by condition. For students in the target condition, their goal was to reach a specific distance. This distance was determined uniquely in each class by calculating the median distance of all initial scooter designs (see Figure 5). This ensured that half of the class would need to travel farther, but half of the class would need to make their scooter travel less distance. On the other hand, for students in the distance condition, their goal was to produce a scooter that would travel the farthest in their class.

Prior to revising their scooters, students are guided through a group collaboration activity. Specifically, students are instructed to meet in groups of 4 (i.e., two pairs) to discuss how to improve their scooters. Pairs are chosen by the experimenter (in collaboration with the teacher) to ensure that one pair’s initial scooter travelled under the
median distance, while the other pair’s scooter travelled further than the median distance. For instance, if the median class distance was 75 cm, one group whose scooter travelled 50 cm could be paired with another group whose scooter travelled 125 cm. In the distance condition this establishes a peer-mentoring situation in which one group with a better initial design can advise the other pair on how to increase distance. In the target condition, neither group would necessarily have a “better” design for the median target goal, because even the group whose initial scooter traveled further would need to revise in order to limit distance.

Figure 5. Testing “lanes” set up for target condition

Reflection. Finally, after completing the redesign and testing, students reflect on the process of revising the scooter and their outcomes. Students are prompted to reflect on the role of energy in their designs and integrate energy concepts with practical issues of design, as well as make final suggestions for future models.

Procedure

We administered a pretest to all students before they began the lesson. Students performed the pretest individually on a set of classroom laptops for approximately 30 minutes. Following the pretest, students were assigned to workgroup pairs based upon teacher discretion (i.e., the teacher assigned partners based on previous classroom projects). Whole classes were assigned randomly to experimental condition. Teacher A had two target condition classes and two distance condition classes, while teacher B had three target condition classes and two distance condition classes. After the students completed the Scooters activity they performed the posttest individually. The posttest took approximately 30 minutes. The entire length of the study from pretest through posttest was 10 days.

Coding and analysis

Pretest and posttest items were scored according to a knowledge integration rubric (scores 1 – 5), which rewards linking ideas into a coherent narrative (Liu, Lee, & Linn, 2011). Lower scores represent incomplete or nonnormative ideas (e.g., I don’t know (1), “everything on the graph changed because if you change one thing on the car it will change everything on the graph” (2)), while higher scores indicate normative ideas that are linked together (e.g., “The potential energy didn’t change because less friction doesn’t have more energy stored” (4, representing one complete link between two ideas). This approach has been used in previous graphing applications (Vitale et al., 2016) to emphasize links between narrative elements of the item (e.g., the “speed”) and spatial elements (i.e., relative height of kinetic energy in a stacked bar graph).

Log data during virtual model exploration was collected with the WISE system. Each time a student clicked on a virtual model, the action was recorded. The “click events” were processed and analysed.

Results

We began this study with three research questions focusing on overall learning gains, learning differences by condition, and qualitative differences in evidence supplied in each condition. We address each in turn.
Learning of energy concepts

To determine whether the Scooters unit effectively teaches energy concepts, we compared scores on all three repeated pre- and posttest items. Using paired, two-tailed t-tests, we found significant increases from pretest to posttest on all three items, Conservation [pretest: M = 2.10, SD = 0.46; posttest: M = 2.32, SD = 0.77; t(224) = 5.1, p < .001, d = 0.35], Revision [pretest: M = 2.32, SD = 0.70; posttest: M = 2.64, SD = 0.84; t(193) = 5.0, p < .001, d = 0.41], and Comparison [pretest: M = 2.66, SD = 1.03; posttest: M = 3.22, SD = 1.03; t(162) = 6.4, p < .001, d = 0.51].

Differences in exploration strategies by condition

The logged data from the virtual model further suggests the target condition better supported students to distinguish relevant variables. Students had the opportunity to interact with the virtual model at three different time points. The following results reflect the students’ first virtual model experience. Students viewed one of two versions of the model: distance or target (Figures 2 and 3). In the distance model, students are told to explore the model. In the target model, students are told to try to hit the target that appears along the scooter’s track. Once students hit the target, the target moves randomly to another position. On average, students ran the model 10.99 (SD = 8.79) times, suggesting that students were actively engaged. Students in the target condition conducted significantly more trials in the model than students in the distance condition (mean target = 14.38, mean distance = 7.55, t(95) = 4.91, p < .001). Further, students in the target condition conducted more efficient trials. In our model, wheel radius did not affect how far the scooter travelled. Therefore, if students understood this, there would be little reason to continue to alter this feature. Students in the target condition altered the wheel radius significantly less than students in the distance condition (target: 24% of runs; distance: 49% of runs; t(133) = 3.77, p < .001), suggesting that more students in the target condition distinguished the impact of wheel radius relative to the other variables on motion. Adjustments to mass and friction did not vary between conditions.

Learning differences by condition

To address which condition generated stronger learning outcomes, we conducted a mixed effects regression with random effects for student and fixed effects for condition, teacher, and pretest total score. A mixed effects model allows us to utilize results from each of the three assessment items to determine the overall impact of each of our predictors (pretest, teacher, condition). Results from the mixed effects regression can be found in Table 2. The results show that, controlling for teacher and pretest score, there is only a trend for condition, suggesting that the target condition was positively related to students’ performance at posttest (β = 0.11, p = .14).

<table>
<thead>
<tr>
<th>Table 2. Regression data for the effect of condition on posttest scores, controlling for teacher and pretest score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Target Condition</td>
</tr>
<tr>
<td>Teacher 1</td>
</tr>
<tr>
<td>Pretest Score</td>
</tr>
</tbody>
</table>

Differences in evidence supplied by condition

Our last research question asks if condition affects how students relate the information from the computer model to building a physical scooter. We begin by looking at the posttest item Consultant, because it provided an appropriate opportunity for students to supply evidence from their instructional experience to justify their responses. (This item appeared at posttest only, and is not included in the analyses above.) This question required students to list as many changes as they could to help improve Lorena’s car, which had stopped in the middle of the track. There was no difference between conditions for the number of changes students suggested (target: M = 2.44; distance: M = 2.36; t(227) = 0.55, p > .1). However, students did differ in terms of the types of changes offered (see Table 3 for examples). Students in the target condition consistently suggested improvements to Lorena’s car that were based on the model. For example, students were more likely to suggest that Lorena increase the air in the balloon (target: 75% of 244 students; distance: 61%; χ²(1) = 4.80, p = .03), reduce the mass of the scooter (target: 45%; distance: 32%; χ²(1) = 3.77, p = .05), and marginally more likely to suggest decreasing friction (target: 59%; distance: 46%; χ²(1) = 3.22, p = .07). Students suggested adjustments to the car’s wheels (an incorrect answer based on the model) at equal rates (target: 31%, distance: 30%; χ²(1) = 0, p > .1).
While students in the target condition appeared to focus on the virtual model more often, students in the distance condition were more likely to suggest practical changes to the car that were based on their own building experiences. For example, students in the distance condition more frequently suggested that Lorena makes sure her wheels are straight, uses a straw over her axle (a practical solution for decreasing friction that was not shown in the model), or adjusts the amount of tape on her scooter. On average, 28% of target students and 44% of distance students offered suggestions related to practical issues, as opposed to the mechanistic variables influencing energy transformation. The difference between the two conditions was significant ($\chi^2(1) = 5.38, p = .02$). This suggests that the target condition may be more effective in helping students to distinguish variables and their relationship to both the underlying science principles and the physical design than the distance condition.

<table>
<thead>
<tr>
<th>Target condition</th>
<th>Distance condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I would recommend that she make the wheels of the car larger, reduce friction between that axle of the wheels and the body of the car, and inflate the balloon more.</td>
<td>You could change it from a balloon to a rubber band or you could add a straw.</td>
</tr>
<tr>
<td>2. To increase the distance of a Newton’s scooter, Lorena can: Blow up the balloon more, decrease the friction, decrease the mass, or decrease the wheel size.</td>
<td>She can fill up her balloon with more air in the beginning so it can move farther, she can choose a stretchier balloon, and she can find a way to make the air come out slower (she could use a straw).</td>
</tr>
<tr>
<td>3. I think she could blow up the ballon more so there is more potential energy that can later on be turned to kinetic and she could make the wheels slightly bigger.</td>
<td>Change the placement of the balloon. Make the tires more straight.</td>
</tr>
<tr>
<td>4. 1. You could reduce friction 2. Put less mass 3. Fill the balloon up more 4. The wheels are bigger</td>
<td>To make it go further maybe add more friction on to the wheels by taping the sides around the wheel. You can tape more with the cd on to the skewer.</td>
</tr>
</tbody>
</table>

**Discussion**

Designing engaging, rigorous, and hands-on inquiry projects requires trial and refinement (Edelson et al., 1999). Designing multi-faceted inquiry projects that address both core disciplinary concepts and scientific practices is particularly challenging. However, such activities can integrate science practices and content knowledge, as well as engage students in ways that are likely to make a lasting impact. Thus, developing approaches to support student design projects is an essential goal of educational research.

In this study we found some evidence for engaging students in exploring trade-offs by focusing students on using their scooter to hit a target. The target design constraint led to more exploration of trade-offs than did the distance design constraint. This is consistent with findings of Schauble, Klopfer, and Raghavan (1991). They found that when an activity prompts students to control variables, students explore variables carefully and deliberately. On the other hand, when the activity prompts students to make a boat go as far or as fast as possible, they chose experiments focused on the goal, rather than systematically controlling variables that are not essential to the solution. In the present study, the opportunity to integrate experimentation with the virtual model appears to have engaged both the target and the distance condition in exploring the variables, suggesting that the specific design features of the activity involving distance are important.

Combining a virtual model and a physical design activity requires careful design and analysis of conditions to ensure that the students succeed. This study highlights both the unique affordances of physical models and virtual models, and the challenges of combining the two formats (de Jong, Linn, & Zacharia, 2013). For example, Zacharia, Olympiou, and Papaevripidou (2008) found that students who learned about heat and temperature through a combination of a physical activity and a virtual model made greater gains than students who learned using physical activities only.

One of the successes of combining physical activities and virtual models can be seen in the use of the virtual model in the Consultant question. Both groups engaged with the virtual activity. Students in the target condition were more likely than those in the distance condition to draw on the virtual model when providing ideas to the Consultant question. In previous pilot iterations of the study, without the target condition, students infrequently
used the model as evidence when discussing their ideas about how scooters work. Students appeared to think only of the hands-on activity and did not view the virtual-physical activities as integrated. By providing suggestions from the virtual model, students in the target condition showed that they are beginning to consider the virtual model as providing evidence that can be applied outside the model context. This suggests that students in the target condition recognized the value of the virtual model for distinguishing among variables to hit a target. More research is needed to validate this conjecture.

The use of both virtual and physical approaches generates challenges for teachers who may worry that building physical models could distract from learning science principles. Many teachers prefer to limit hands-on projects to extracurricular activities. To make physical model building effective, we clearly specified the materials and procedures for building scooters, enabling teachers to prepare for construction. Furthermore, by guiding students’ initial designs in WISE, we ensured that students had similar experiences. Using WISE, students could make plans, describe their designs, interpret evidence, and link their results to energy principles. Teachers could use the dashboard to evaluate student thinking as it developed. These supports enabled efficient implementation of the activities, consistent with the finding that all students made clear learning gains on each of the posttest items focused on energy.

This study also illustrates the importance of determining constraints carefully when designing engineering projects. Similar issues emerge in research on games. Games impose design constraints to lengthen the experience, induce creativity, and add competition. Games can improve learning by focusing constraints on science-related phenomena and providing appropriate feedback. Recognizing the relevancy for education, a number of researchers have suggested that more game-like features should be introduced into educational applications while also cautioning designers to explore the addition of these features carefully to avoid unintended consequences (see Clark, Tanner-Smith, & Killingsworth, 2016, for overview and meta-analysis). For example, in a learning activity with a game-like simulation, Miller, Lehman, and Koedinger (1999) found that students learned relevant physics concepts (related to electromagnetism) better when their actions in the simulation were limited to specific options than when they could engage in more open-ended exploration. In this case, specifying the options compelled learners to engage challenging problems that could be avoided in open-ended navigation. This illustrates how successful designs for constraints can direct student attention to key inquiry activities.

Even with the support of online tools, there is potential for students to focus on superficial elements of their engineering designs and overlook critical features. Ensuring that the design elicits scientific thinking is an important task in instructional design. In this study, we hypothesized that the target condition (compared to the distance condition) would encourage linking the virtual model exploration and physical building tasks and result in more careful attention to the relations between scooter features and underlying energy concepts. We found evidence in the Consultant item that students in the target condition drew more on the virtual model than students in the distance condition (i.e., they were more likely to refer to the model in their responses). Nevertheless, the impact on learning energy concepts was similar. Although the target condition did increase emphasis on trade-offs, students may not have fully analyzed the model to recognize relationships such as between adding friction and its impact on transfer of kinetic energy to thermal energy. In future research we will explore ways to encourage students to focus and make predictions directly from information on the energy graphs.

Despite a lack of clear evidence to differentiate learning of energy concepts by condition, the results suggest that the design constraint approach helps clarify the interactions that arise when students combine virtual and physical designs. Determining how best to design activities with constraints remains a process of trial and refinement.

**Implications for teaching and instructional design**

While many teachers feel pressure to cover a wide range of topics within a short time-span, educational standards highlighting scientific practices (e.g., NGSS Lead States, 2013) offer an opportunity to engage students in exploring complex science concepts. Ensuring that hands-on projects meet content requirements requires critical analysis of the design activity. In designing effective activities, it is essential to link activities to science principles. Although web-based systems are one tool for addressing this question, teachers may supplement hands-on inquiry with a diverse array of activities, such as journal writing, drawing, diagramming, graph construction, etc. Each of these activities could reinforce the scientific mechanisms and build upon students’
ideas (Linn & Eylon, 2011). This study suggests the value of studying design constraints and refining them to help students think creatively and engage in knowledge integration.

Acknowledgements

This material is based upon work supported by the National Science Foundation under NSF Projects 1418423 and 1451604. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References


Design of a Three-Dimensional Cognitive Mapping Approach to Support Inquiry Learning

Juanjuan Chen¹, Minhong Wang¹*, Chris Dede² and Tina A. Grotzer²
¹KM&EL Lab, Faculty of Education, The University of Hong Kong, Hong Kong // ²Graduate School of Education, Harvard University, MA, USA // jchen12@connect.hku.hk // magwang@hku.hk // chris_dede@harvard.edu // tina_grotzer@harvard.edu

*Corresponding author

ABSTRACT

The use of external representations has the potential to facilitate inquiry learning, especially hypothesis generation and reasoning, which typically present difficulties for students. This study describes a novel three-dimensional cognitive mapping (3DCM) approach that supports inquiry learning by allowing learners to combine the information on a problem, the subject knowledge, and the hypothesizing and reasoning process involved in the exploration in a single image. The study also investigates the influences of the 3DCM approach on knowledge achievement and learner perceptions within an online inquiry-learning context. Forty-eight 11th-grade students used 3DCM to complete an inquiry task. Data were collected from multiple sources, including pre- and post-tests, questionnaires, and semi-structured interviews. The results revealed that the students showed high levels of academic achievement, positive attitudes toward inquiry learning, low levels of anxiety, and medium levels of confidence. A post-hoc test indicated that the students at a low academic level had acquired significantly more knowledge than either the high-level or medium-level students, thus narrowing the academic gap between low-level, medium-level, and high-level students. In addition, the participants’ attitudes and degree of confidence were found to be positively related to their inquiry skills, such as hypothesis generation and reasoning. The implications of the study and directions for future work are also discussed.

Keywords

Inquiry learning, Reasoning, External representation, Cognitive mapping, Science education

Introduction

Education today emphasizes students’ active involvement in learning, especially through inquiry and problem-solving experiences (Bransford, Brown, & Cocking, 1999). Originating in practices of scientific inquiry, inquiry learning engages students in exploring phenomena or problems by asking questions, collecting and interpreting data, constructing evidence-based arguments, and forming conclusions (Lazonder & Harmsen, 2016). Through inquiry activities, students acquire subject knowledge, develop discipline-related practical experience and reasoning skills (Hmelo-Silver, Duncan, & Chinn, 2007), and enhance motivation (Phielix, Prins, & Kirschner, 2010). The beneficial effects of inquiry learning have been reported on students at various ability levels, from low to high (Zohar & Dori, 2003).

In traditional inquiry-learning contexts, learners often interact with objects in the real world. For example, during physics experiments, they conduct physical hands-on investigations and interact directly with the material world. Information and communication technologies (ICT) are used increasingly in inquiry learning to present problem contexts in vivid and interactive formats. An important affordance of technology-supported inquiry learning is its capacity to highlight salient information and remove irrelevant details, facilitating the interpretation of the phenomena. Furthermore, such learning environments enable students to conduct experiments to investigate unobservable phenomena (e.g., the travelling of light rays). Technology-supported inquiry learning has been demonstrated to be more efficient than traditional inquiry because it provides simulated results instantaneously and enables students to gather more information in the same amount of time (De Jong, Linn, & Zacharia, 2013). Examples of technology-supported inquiry learning include the Web-based Inquiry Science Environment (WISE) (Linn & Slotta, 2000) and immersive learning environments (Dede, 2009).

However, whether working in traditional or technology-enabled contexts, students often experience difficulties in regulating the inquiry process and engaging in fruitful inquiry learning. The inquiry process often involves iterative cycles of gathering information through observation or experiments, generating hypotheses, reasoning based on the collected information, and drawing conclusions (Kuhn, Black, Keselman, & Kaplan, 2000). Generating hypotheses entails formulating ideas about the relationships between variables (Gijlers & de Jong, 2013). Scientific reasoning involves questioning initial premises, seeking evidence that confirms or contradicts the hypotheses, revising initial ideas, and considering alternative hypotheses (Zeineddin & Abd-El-Khalick, 2010). Many students do not know how to formulate hypotheses (de Jong & van Joolingen, 1998). In addition, students' reasoning ability may be inadequate (Zeineddin & Abd-El-Khalick, 2010). For example, some students...
find it difficult to connect evidence with claims or to reason using intertwined variables (Kamarainen, Metcalf, Grotzer, & Dede, 2014). Studies have also indicated that some students are unable to adapt or revise an initial hypothesis in the presence of conflicting evidence (Kuhn et al., 2000). Furthermore, some students rarely reflect on prior conceptions and readily dismiss contradictory information (Zeineddin & Abd-El-Khalick, 2010).

In light of these findings, it has been argued that it is urgently necessary to support inquiry learning by guiding students through the complex inquiry process and helping them to become accomplished problem-solvers (Kirschner, Sweller, & Clark, 2006). Hmelo-Silver et al. (2007) argued that effective problem-based learning involves significant, built-in supporting structure. More and more researchers are investigating a continuum of forms of support and guidance for inquiry learning. To facilitate the complex inquiry process, some strategies such as scripts, prompts, and hints (Kollar, Fischer, & Slotta, 2007) have been provided to facilitate the inquiry process (i.e., what to do next or how to do it). When developing supportive frameworks, however, it is important to avoid undermining the open-endedness of the task and individual endeavour. Accordingly, additional forms of cognitive supports based on cognitive maps have been explored and shown their promising effects, such as concept maps (Gijlers & de Jong, 2013), causal maps (Slof, Erkens, Kirschner, Janssen, & Jaspers, 2012), evidence maps (Suthers, Vatrapu, Medina, Joseph, & Dwyer, 2008), and integrated cognitive maps representing the problem-solving process and the underlying knowledge (Wang, Wu, Kinshuk, Chen, & Spector, 2013; Yuan, Wang, Kushniruk, & Peng, 2016). A cognitive map is an external representation of cognitive structures and processes. Cognitive maps are deemed to effectively engage students and to foster high-order thinking and meaningful learning in complex situations (Jonassen, 2005).

This study outlines a novel three-dimensional cognitive mapping (3DCM) approach that supports inquiry learning by allowing learners to combine, in a single image, information on a problem, subject knowledge (key concepts and their relationships), and the process of hypothesizing and reasoning involved in exploring the problem. The premise that underlies the design is that externalizing the complex aspects of an inquiry task makes inquiry learning more accessible to learners (Janssen, Erkens, Kirschner, & Kanselaar, 2010).

The aim of this study was to explore the influences of the 3DCM approach on inquiry learning in an online environment. The benefits of constructing external representations to support inquiry learning can be explained on the cognitive (e.g., academic achievement), metacognitive, and social dimensions (Toth, Suthers, & Lesgold, 2002). In addition to their academic achievement, student motivational and emotional experiences were investigated with reference to their attitudes towards inquiry learning, perceived inquiry skills, anxiety level, and confidence level. These experiences have been shown to be significantly related to learning achievements (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011), although they have rarely been investigated thoroughly in technology-enhanced learning environments (Järvenoja & Järvelä, 2005).

In addition, the study examined the impacts of 3DCM approach on low-ability, medium-ability, and high-ability students, respectively, because it is critical to understand the differential effects of scaffolding or support on students of dissimilar academic levels (Veermans & Järvelä, 2004). It is logical to hypothesize that the impact of a specific instructional method and scaffolding/support will vary by academic level because high-level and low-level students differ in what they can and cannot do. However, few studies have investigated the impacts of scaffolding/support on students of different academic levels (Belland, Glazewski, & Richardson, 2011).

**Literature review**

**External representation**

External representations are graphical or diagrammatic representations of knowledge or information in the form of maps, diagrams, tables, or pictures (Cox, 1999; Toth et al., 2002). They can be constructed by the students themselves during learning sessions created beforehand by teachers, or taken from textbooks. Learners’ active construction of maps related to the solution of a given problem has been found to promote learning in inquiry and problem-solving contexts in the dimensions of cognition, metacognition, and socio-emotional development (Janssen et al., 2010; Suthers et al., 2008; Wang et al., 2013). In the cognitive dimension, constructing external representations assists problem solving by clarifying learners’ thinking, helping them to re-order information and draw inferences, sustaining their focus on the construction of knowledge, and helping to reduce their cognitive load (Cox, 1999). Metacognitively, the use of external representations tracks the progress of reasoning and directs attention to the unsolved part of the problem. In the social aspect, external representation serves as an discussion anchor that coordinates the discourse between peers by making one’s thinking visible to others (Schwendimann & Linn, 2016).
Among various external representations, concept maps have been widely used in learning. Although most studies on concept mapping have used it as a conceptual learning tool (Schwendimann, 2015), several researchers have deployed this type of representation to facilitate inquiry learning and group tasks. For example, Gijlerts and de Jong (2013) found that students who constructed concept maps performed better on knowledge tests in a simulation-based inquiry learning program. Collaborative concept mapping was found to be more effective in supporting group interaction in the situation of concept-oriented task than in that of design-oriented task (Wang, Cheng, Chen, Mercer, & Kirschner, 2017), and role assignment in concept mapping mediated small group learning improved socio-emotional experiences (Cheng, Wang, & Mercer, 2014).

Other popularly used external representations include causal maps, evidence maps, and integrated cognitive maps. Causal mapping is a form of concept mapping used to represent relationships of cause and effect. Slof et al. (2012) found that students who created causal maps offered better justification for their solutions than those who did not create causal maps. Evidence maps that link evidence with claims or hypotheses were found to support hypothetic reasoning (Suthers et al., 2008). Wu and Wang (2012) proposed a computer-based cognitive-mapping approach that enabled students to externalize the hypothetic reasoning process and the underlying knowledge in an integrated cognitive map when they work with diagnostic problems. This approach showed promising effects on diagnostic problem-solving (Wu, Wang, Grotzer, Liu, & Johnson, 2016).

External representations have been shown to promote student learning, especially with the potential to benefit low academic level students (Liu, Chen, & Chang, 2010; Schnotz, 2002). For example, O'Donnell, Dansereau, and Hall (2002) showed that low academic level students benefited the most from knowledge mapping and recalled more knowledge than students of other academic levels. Moreover, in the study on a graphical representation (i.e., proof tree), Wong, Yin, Yang, and Cheng (2011) found that medium-level students reported the most enjoyment.

**Learners’ motivation and emotion**

Learners’ motivation and emotion have been shown to significantly relate to their learning achievements (Pekrun et al., 2011). “Motivation” refers to the psychological characteristics that drive students to persist in working toward their learning goals (Muilenburg & Berge, 2005), mainly including beliefs and attitudes such as confidence, self-competence, and interest. Of these, confidence appears to be particularly salient to learning, especially self-regulated learning, because it influences the degree to which learners engage and persevere when facing challenging tasks (Jones & Issroff, 2005).

Emotion encompasses learners’ positive and negative reactions to teachers, classmates, schools, and instructional methods (Fredricks, Blumenfeld, & Paris, 2004). Anxiety, boredom, and frustration are typical negative emotions; curiosity, enjoyment, and pride are common positive emotions. Negative academic emotions can impede cognitive processes, whereas positive ones can foster learning (Pekrun et al., 2011). If a learning task is too complex, students may feel frustrated and find it difficult to figure out what to do, which in turn increases their anxiety (Schutz & DeCuur, 2002). Various studies have suggested that the use of cognitive tools would be effective in promoting motivation and/or positive emotions. For example, the use of knowledge maps was shown to enhance motivation and positive attitudes (Sung & Hwang, 2013).

**Design of a 3DCM-supported learning environment**

**Three-dimensional cognitive mapping**

Because inquiry and problem-solving tasks usually combine various kinds of information, data, concepts, and relationships, many students find it cognitively demanding to integrate problem data with subject knowledge and to reason using intricately intertwined data. Therefore, more research is needed to investigate methods of externalizing complex cognitive processes to achieve the desired learning outcomes. We studied the use of a *Three-dimensional Cognitive Mapping (3DCM)* to support inquiry learning by allowing learners to combine in a single image problem information, subject knowledge (key concepts and their relationships), and the processes of hypothesizing and reasoning.

An example of 3DCM is illustrated in Figure 1. The figure shows an expressive representation of a problem with three parts: a concept map, a data table, and a reasoning map. The concept map illustrates the concepts...
underlying the problem and the relationships between these concepts. The data table records the problem information, reflected as a set of key variables and their changes over an observation period. The reasoning map represents the evidential relationships between the hypotheses and the data or subject knowledge. In the reasoning map, each hypothesis is supported (“for”) or rejected (“against”) by evidence from the data or subject knowledge. To examine the root cause of the problem, the hypothesis is further explained by other hypotheses that explicate deeper causes of the problem. Learners draw the reasoning map while observing the concept map and data table.

![Figure 1. The Three-dimensional cognitive mapping](image)

**Question:** Why has the number of wolf on the grassland increased a lot?

**Concept Map**

- **Wolf**
- **Deer**
- **Mice**
- **Owl**
- **Grass**

**Data Table**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolf</td>
<td>The number of wolf was stable before last year, increased a lot in this summer.</td>
</tr>
<tr>
<td>Deer</td>
<td>The number of deer has not changed too much these two years.</td>
</tr>
<tr>
<td>Mice</td>
<td>The number of mice increased at the end of last year, but decreases a little this year and becomes about the same as that of the beginning of last year.</td>
</tr>
<tr>
<td>Grass</td>
<td>The grass grew and flourished last summer.</td>
</tr>
</tbody>
</table>

**Reasoning Map**

- **H1:** the increase of wolf was caused by increase of deer
- **H2.1:** the increase of rabbits was related to decrease of mice
- **H2.2:** the increase of rabbits was due to flourishing grass
- **H2:** the increase of wolf may be due to the increase of rabbits

**Legend:**

- **Hypothesis**
- **Data evidence**
- **Knowledge evidence**
- **Supporting evidence**
- **Rejecting evidence**
- **Hypothesis that is rejected**
- **Numbering of hypothesis**

**Learning environment and materials**

To support situated learning, an online learning environment that presents a pollution problem within an ecosystem was designed and implemented. It consists of two major modules: problem context and learning support.

The *problem-context* module presents an authentic pollution problem in a pond ecosystem as a learning task that requires consideration of multiple perspectives and the use of evidence to reach an adequate solution. The problem-context was based on the EcoMUVE curriculum (Kamarainen, Metcalf, Grotzer, & Dede, 2015), in which students explore a virtual pond and the surrounding watershed, observe simulated organisms for a number of virtual “days,” and collect relevant data in order to investigate why many of the fish had died overnight. The module contains *information-collection* and *data-observation* sub-modules. As shown in Figure 2, the sub-module of *information-collection* provides a rich context, such as descriptions and visualizations of the surroundings and background information on the pond ecosystem. Users select a specific date from a dropdown list of dates to observe and collect information that provides tacit clues or hints and guides students in observing the phenomena.
The sub-module of *data-observation* shown in Figure 3 facilitates problem exploration by presenting graphs of the data on the key variables and their changes over the observation period. For example, students could observe data on water conditions (e.g., water temperature, turbidity, dissolved oxygen), weather conditions (e.g., air temperature, wind speed), and the population of various organisms (e.g., bacteria, algae, bass) in the pond ecosystem. The module also enables learners to construct queries by selecting variables to facilitate the analysis and comparison of variables.

**Figure 2. Information collection**

**Figure 3. Data observation**

The *learning support* module affords some useful learning guidelines. First, it provides domain-specific knowledge about ecosystems and ecological processes, such as the food web and photosynthesis. Second, the module introduces the basic skills required for scientific inquiry and the steps required for scientific inquiry and the fundamental steps involved, such as hypothesis formation and evidence-based reasoning. Third, the module provides general instructions on building a 3DCM, and conducting group discussions, along with an example to demonstrate how to use the learning system.

**Research questions**

This study’s research questions are stated as follows:

RQ1. What are the influences of 3DCM on students’ learning outcomes (i.e., knowledge achievement, inquiry skills, attitudes, anxiety level, and confidence level)?

RQ2. What are the influences of 3DCM on the performance of students of different academic levels?

RQ3. What are the relationships between different learning outcomes?
Methodology

Research design

A study was conducted to evaluate the feasibility and influences of the proposed 3DCM approach to support inquiry learning in an online learning environment. The students used the 3DCM to facilitate the inquiry process. They were required to work in small groups to complete the same learning task: exploring a fish die-off problem in a biology course. The students’ learning outcomes were measured.

Participants

Forty-eight students from one 11th grade high school class were recruited to participate in the experiment. The sample comprised 24 male and 24 female students, with an average age of 17 (range: 16 to 18). They were classified into three categories of academic ability according to their pre-test scores: high, medium, and low, with each category having 16 students. Students were then randomly divided into 16 small groups of 3 (i.e., one high-level, one medium-level, and one low-level student), which is a typical method of heterogeneous ability grouping for group learning (e.g., Phielix et al., 2010) and has shown its effectiveness in fostering learning (Lou et al., 1996).

Learning task

The learning task required the students to perform causal reasoning and to construct logical and scientific explanations for a fish die-off problem; that is, why so many large fish in a pond ecosystem had suddenly died. The students could freely play with the online environment to collect relevant information and observe changes in each variable over time. They discussed and solved problem in small groups by evaluating and compiling the collected information, formulating hypotheses, and reasoning. They were also asked to create a 3DCM to assist their inquiry. Finally, every group was required to submit an inquiry report to present their solutions, which should include their hypotheses, reasoning, and conclusions.

Procedure

The experiment was conducted in six 45-minute sessions over 2 weeks. During the first session, consent forms were signed by the participants. A pre-test questionnaire was administered individually, and small groups were formed.

At the beginning of the second session, the researcher provided a 20-minute introduction and demonstration on how to perform inquiry learning using the online system in the school’s computer laboratory. The principles of social interaction, the method of constructing a 3DCM, and the supporting materials were also briefly introduced.

After the demonstration, students embarked on the learning task in the computer laboratory at school, with group members sitting together and each member accessing a networked computer. After viewing the relevant information and data individually, each student might have developed some ideas or initial hypothesis about the problem. They then began group discussion and collaboration; each group member was encouraged to share personal ideas within the group and to respond to challenges from group mates. To support or reject their initial hypothesis, they collaboratively made observations, collected information, and formulated hypotheses from multiple perspectives by brainstorming ideas based on the compiled evidence (i.e., information and data). Finally, the group had to reach an agreement on the best explanations of the problem. At the end of the fifth session, each group was asked to submit an inquiry report. Next, a semi-structured written interview was conducted to investigate the students’ perceptions of the 3DCM approach by asking them to write the responses on paper. Finally, in the sixth session, the participants were asked to individually complete a knowledge test and a questionnaire survey.
Measures and instruments

Pre-test questionnaire

The pre-test questionnaire elicited information on students’ gender, age, and self-assessment of their computer skills (“Please describe your computer skills: 1 = Very poor; 2 = Poor; 3 = Neither poor nor good; 4 = Good; 5 = Very good.”)

Knowledge test

Both pre- and post-knowledge tests were administered to assess the students’ knowledge of the learning subject. Each test comprised multiple-choice questions, fill-in-the-blank questions, and short-answer essay questions to assess students’ knowledge of ecosystems, photosynthesis, respiration, and decomposition. The post-test was directly related to the learning task (i.e., a problem within a pond ecosystem). The highest possible score on the knowledge test was 100. The pre- and post-test were at the same level of difficulty. Both tests were designed collaboratively by the teacher and the researcher together. The teacher and another biology instructor expert evaluated the two tests to determine their difficulty level and to confirm their content validity.

Post-test questionnaire

A post-test questionnaire was administered to measure students’ attitudes toward inquiry learning, perceived inquiry skills, anxiety level, and confidence level. Responses were given on a 5-point Likert-scale ranging from 1 (strongly disagree) to 5 (strongly agree). Six items measuring attitudes were drawn from Chu, Hwang, Tsai, and Tseng (2010), such as, “I think inquiry learning is more interesting than traditional instruction.” Four items measuring perceived inquiry skills used in the learning task were adapted from De Jong (1991), such as, “I can generate testable hypotheses from different perspectives.” The inquiry skills measured were hypothesis generation, reasoning, and conclusion making. The students’ self-reported anxiety and confidence during the learning task were also measured. To measure anxiety, three items were developed based on the work of Pekrun et al. (2011), such as, “I felt anxious while doing this task.” Three items to measure confidence were developed from the work of Keller (2010), such as, “I felt confident during the processes of inquiry learning.” Higher scores represented more positive attitudes, better inquiry skills, greater anxiety, and higher confidence, respectively.

Semi-structured written interview

The interview was conducted to get the students’ perceptions of the 3DCM approach, using the following questions. “Did the 3DCM approach help you to solve the problem?” “If so, how did it help you?” Students were asked to write their responses to the questions on paper.

Data analysis methods

The following methods of data analysis were used.

- To answer RQ1, statistics were obtained for learning outcome measures, including means and standard deviations (SDs), and a paired-sample t-test was conducted to compare the difference between post- and pre-knowledge test scores.
- To answer RQ2, one-way analysis of variance (ANOVA) was performed to evaluate the statistical differences in learning outcomes between students of three different academic levels, with academic level as a between-subject independent variable. Since a statistical difference in knowledge gain was found between three levels of students, we then run Tukey’s post-hoc test to compare each of the three levels to every other level to figure out which specific level of students differed from others.
- To answer RQ3, correlation analysis was run to gain a deeper understanding of the relationships between the measured learning outcomes.
- A thematic content analysis of the interview was performed to identify common themes in students’ responses to the written interview question. It was performed by one of the researchers and a research assistant with expertise in content analysis. Cohen’s Kappa value for inter-rater reliability was .78.
Results

Pre-test

The 48 students’ mean score on the pre-test was 63.19 (SD = 8.49). As shown in Table 1, high-level students were those scoring the top third, with a mean score of 71.94 (SD = 2.77); medium-level students were those scoring the middle third, with a mean score of 64.25 (SD = 2.84); and low-level students were those scoring the bottom third, with a mean score of 53.37 (SD = 4.95). ANOVA result revealed significant differences in prior knowledge between students at the three levels (F (2, 45) = 103.75, p < .001). The mean score for the participants’ self-perceived computer skills was 3.29 (SD = .51), and no significant differences were observed between the students at different academic levels (F (2, 45) = 1.16, p > .05).

RQ1. What are the influences of 3DCM on students’ learning outcomes (i.e., knowledge achievement, inquiry skill, attitude, anxiety level, and confidence level)?

All 48 participants completed the post knowledge test. Yet, only 38 returned their completed questionnaires because the other 10 students were absent during the survey session. As shown in Table 1, the mean score on the post knowledge test (mean = 78.26, SD = 8.69) was higher than that on the pre-test (mean = 63.19, SD = 8.80); this difference was statistically significant (t_{47} = 9.576, p < .001), based on the paired-sample t-test result. This finding indicates a significant improvement in the students’ domain knowledge.

Regarding the reliability of the questionnaire, the Cronbach’s alpha values were .90 for attitude, .83 for inquiry skills, .74 for anxiety, and .72 for confidence, indicating credible internal consistency. The means were 4.15 (SD = .54) for attitude, 4.20 (SD = .49) for inquiry skills, 2.44 (SD = .77) for anxiety, and 3.82 (SD = .61) for confidence, indicating fairly high inquiry skills, positive attitude, low anxiety, and medium confidence. With respect to inquiry skills, the mean scores were 4.18 (SD = .61) for hypothesis generation, 4.13 (SD = .66) for scientific reasoning, and 4.32 (SD = .62) for making conclusions. All of these values were fairly high (exceeding 4 out of 5), suggesting that the students perceived themselves to have performed quite well in all aspects of inquiry processes.

Table 1. Descriptive statistics for learning outcomes and ANOVA results for students at different academic levels

<table>
<thead>
<tr>
<th>Type of learning outcome</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-knowledge test</td>
<td></td>
<td></td>
<td></td>
<td>F (2, 45) = 103.75 (p &lt; .001)</td>
</tr>
<tr>
<td>High</td>
<td>16</td>
<td>71.94</td>
<td>2.77</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>16</td>
<td>64.25</td>
<td>2.84</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>16</td>
<td>53.37</td>
<td>4.95</td>
<td></td>
</tr>
<tr>
<td>Post-knowledge test</td>
<td></td>
<td></td>
<td></td>
<td>F (2, 45) = .59 (p = .56)</td>
</tr>
<tr>
<td>High</td>
<td>16</td>
<td>79.63</td>
<td>8.89</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>16</td>
<td>78.77</td>
<td>8.70</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>16</td>
<td>76.38</td>
<td>8.72</td>
<td></td>
</tr>
<tr>
<td>Attitude</td>
<td></td>
<td></td>
<td></td>
<td>F (2, 35) = .70 (p = .51)</td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>4.03</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>13</td>
<td>4.28</td>
<td>.66</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>13</td>
<td>4.14</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>Inquiry skills</td>
<td></td>
<td></td>
<td></td>
<td>F (2, 35) = .21 (p = .81)</td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>4.13</td>
<td>.56</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>13</td>
<td>4.24</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>13</td>
<td>4.23</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
<td>F (2, 35) = 2.03 (p = .15)</td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>2.42</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>13</td>
<td>2.74</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>13</td>
<td>2.15</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td></td>
<td></td>
<td></td>
<td>F (2, 35) = .15 (p = .86)</td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>3.81</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>13</td>
<td>3.77</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>13</td>
<td>3.90</td>
<td>.53</td>
<td></td>
</tr>
</tbody>
</table>
RQ2. What are the influences of 3DCM on the performance of students of different academic levels?

The mean scores on post-test obtained by the high, middle, and low academic level students (as defined by their pre-test scores) were 79.63 ($SD = 8.89$), 78.77 ($SD = 8.70$), and 76.38 ($SD = 8.72$), respectively, as shown in Table 1. ANOVA result revealed no significant differences in post knowledge test scores between the students of the three levels ($F(2, 45) = .59, SD = .56$). In terms of the knowledge gain (i.e., the difference between post-test and pre-test scores), the mean values were 7.69 ($SD = 8.94$) for the high-level students, 14.52 ($SD = 9.35$) for the medium-level students, and 23 ($SD = 8.91$) for the low-level students. This demonstrates the effectiveness of 3DCM in improving all students’ knowledge regardless of their prior knowledge level. ANOVA result revealed significant difference between the students at the three levels in knowledge gain ($F(2, 45) = 11.44, p < .001$), indicating that academic level had a significant effect on knowledge gain. Post-hoc comparisons using the Tukey HSD test indicated that the low-level students acquired significantly more knowledge than the high- and medium-level students at the $p < .05$ level, respectively. However, no significant difference in knowledge gain was observed between the high- and medium-level students.

Regarding the students’ perceived inquiry skills, attitudes, anxiety level, and confidence level, ANOVA results showed no significant differences between the three levels of students, as demonstrated in Table 1.

RQ3. What are the relationships between different learning outcomes?

As seen in Table 2, both the students’ attitude and their confidence were positively related to their perceived inquiry skills; anxiety and confidence were statistically negatively correlated.

<table>
<thead>
<tr>
<th>Table 2. Pearson’s correlations between different learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Post-test</td>
</tr>
<tr>
<td>Attitude</td>
</tr>
<tr>
<td>Inquiry skills</td>
</tr>
<tr>
<td>Anxiety level</td>
</tr>
<tr>
<td>Confidence level</td>
</tr>
</tbody>
</table>

*Note. “p < .01.

Written interview

The students’ responses to the interview question about the benefits of 3DCM are presented in Table 3. They made 45 different positive comments in total. Twenty-six of the students reported that the 3DCM approach had helped them to think in an organized and logical way. These responses indicate that 3DCM had fulfilled its intended functions.

<table>
<thead>
<tr>
<th>Table 3. Students’ responses to the semi-structured interview question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the 3DCM approach help you to solve the problem? If so, how did it help you?</td>
</tr>
<tr>
<td>Students’ comments</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Data and information collection</td>
</tr>
<tr>
<td>Helping us to visualize the data.</td>
</tr>
<tr>
<td>Facilitating the integration and organization of information.</td>
</tr>
<tr>
<td>Systematic thinking</td>
</tr>
<tr>
<td>Helping us to think in an organized and logical way.</td>
</tr>
<tr>
<td>Covering all relationships (e.g., causal relationship).</td>
</tr>
<tr>
<td>Helping us to find the right direction to solve the problem.</td>
</tr>
<tr>
<td>Hypothesizing</td>
</tr>
<tr>
<td>Enabling us to conduct more in-depth and comprehensive analysis of the causes of fish death.</td>
</tr>
<tr>
<td>Helping us to revise previously formed incorrect hypotheses or ideas.</td>
</tr>
<tr>
<td>Reasoning</td>
</tr>
<tr>
<td>Improving our reasoning abilities.</td>
</tr>
<tr>
<td>The reasoning map clarified the detailed reasoning process.</td>
</tr>
<tr>
<td>Reasoning map guided us step-by-step and helped develop a logical argument progressively.</td>
</tr>
<tr>
<td>Sufficient evidence was included to make the reasoning convincing.</td>
</tr>
<tr>
<td>Helping us to clearly see the relationships between pieces of information/evidence.</td>
</tr>
</tbody>
</table>
**Discussion**

RQ1 addressed the influences of the 3DCM approach on learning outcomes. First, analysis of the pre- and post-knowledge test revealed that the students made significant knowledge gains (measured by the difference in scores between pre- and post-knowledge test), indicating the effectiveness of the approach. This large knowledge gain may be explained by the fact that the designed 3DCM approach effectively supported collaborative knowledge construction. Constructing 3DCM elicited students’ ideas and made students elaborate on them, thus improving their knowledge understanding. The concept map in 3DCM integrated concepts and their relationships, promoting knowledge integration. Research has also suggested that reasoning activities help students to integrate new information with prior knowledge to develop deep, contextualized, and applicable knowledge (Keselman, Kaufman, Kramer, & Patel, 2007; Keys, 2000). Moreover, the 3DCM approach helped students to reflect on their existing knowledge and identify knowledge gaps. The benefits of external representations in facilitating these elaborative, integrative, and reflective processes have also been highlighted by Suthers et al. (2008).

Second, the high mean score obtained for perceived inquiry skills revealed that the students perceived the tool as useful in fostering their problem-solving skills. This was consistent with the findings of previous studies (Janssen et al., 2010; Sung & Hwang, 2013). More specifically, the students reported fairly high levels of skills for generating reasonable hypotheses, reasoning, and drawing conclusions. This result was validated by the interview data. Most of the interviewees reported that 3DCM had enabled them to develop a logical set of hypotheses, and to review and revise previously formed hypotheses or ideas. This confirmed the belief that well-designed external representation helps students to refine their ideas (Cox, 1999).

Regarding reasoning, the reasoning map focused the students’ attention on developing logical justification by finding confirming and disconfirming evidence and linking evidence with their hypothesis. In addition, it helped the learners to clearly see the relationships between different pieces of information/evidence. Moreover, some students said that drawing a data table helped them to notice data that might be omitted. In addition, some students reported that constructing the maps helped them to draw conclusions.

Third, the students’ attitudes toward inquiry learning were quite positive. This was also evident from their interview data. In addition, during the free discussion between the researcher and the participants after the experiment, most students revealed that this inquiry learning approach was very interesting, stimulated their passions in exploring the problem, and made them become active in learning. Sung and Hwang (2013) also showed that the use of external representations in inquiry learning led to a positive attitude.

Fourth, the students’ confidence failed to reach a high level (i.e., > 4) in the inquiry process (mean = 3.82), due to several possible reasons. Participants rarely had prior experiences with such a learning approach. They had been accustomed to passive learning, while the proposed active learning approach was totally new to them. Unfamiliarity with creating 3DCM might reduce their confidence. Moreover, the open-ended task presented to them seemed difficult and challenging. Task is often a source of motivation and emotion (Wosnitza & Volet, 2005). As reflected by some participants, it was their first time solving this kind of problem. These factors may also explain the result for anxiety (mean = 2.44). In addition, technical problems may have contributed to the students’ anxiety; for example, one student said that he needed more time to become familiar with the system. Unfamiliarity with technology was also reported as a cause of anxiety by Wosnitza and Volet (2005). A few students reported Internet connection problems, causing a lack of time in conducting the inquiry activities. Time pressure, also found in other research, may also trigger anxiety (Janssen et al., 2010).

RQ2 explored the influences of 3DCM on students of different academic levels. We found that the students of the three academic levels differed significantly in terms of knowledge gain. Specifically, the low-level students benefited significantly more from the 3DCM approach than the students at the other two levels. This result was consistent with the findings of previous studies, which have shown that cognitive tools or external representations are especially effective in improving the learning achievement of low-level students and

---

### Note

Count = number of students who made similar comments.

<table>
<thead>
<tr>
<th>Source of motivation and emotion</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drawing a data table helped us notice the data that might be missed or omitted.</td>
<td>1</td>
</tr>
</tbody>
</table>

**Communication**

| Helping others to easily understand my ideas. | 1 |
| Recording the key viewpoints. | 2 |

---

### Table

<table>
<thead>
<tr>
<th>Source of motivation and emotion</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drawing a data table helped us notice the data that might be missed or omitted.</td>
<td>1</td>
</tr>
</tbody>
</table>

**Communication**

| Helping others to easily understand my ideas. | 1 |
| Recording the key viewpoints. | 2 |

---

### Note

Count = number of students who made similar comments.
narrowing the gap between low and high academic level students (Belland et al., 2011; Liu et al., 2010; O’Donnell et al., 2002). The students classified as low-level achievers based upon their pre-test scores scored higher than those classified as high achievers based upon their pre-test scores (Dori, Tal, & Tsushu, 2003). Low-level students may be more susceptible to distraction (Armbruster & Anderson, 1980); building external representations can help focus their attentions on relevant learning content (Guastello, Beasley, & Sinatra, 2000). In addition, low-level students in comparison with their high-level peers might need extra support in constructing mental models, which suggests the usefulness of visual representations (Schnotz, 2002).

Additionally, there were no significant differences in learning attitudes, perceived inquiry skills, anxiety level, or confidence level among the students of different academic levels. Similar findings have been reported elsewhere (Liu, 2004). This indicated that all students displayed fairly high levels of learning confidence and perceived inquiry skills when using the 3DCM approach. Although many educators avoid teaching high-order thinking skills to low-level students because they believe that such students are incapable of performing tasks that require high-order thinking (Zohar & Dori, 2003), the findings of the current study suggest that, if taught properly, low-level students can acquire high-order thinking skills as well as high-level students, in line with the findings of Perkins and Grotzer (2005).

RQ3 examined the correlations between different learning outcomes. Both the students’ attitudes and their confidence were positively related to their inquiry skills. Unsurprisingly, it is easy to understand that more positive attitudes toward learning and greater confidence yielded better inquiry performance (Eseryel, Law, Ifenthaler, Ge, & Miller, 2013). The negative correlation between confidence and anxiety was also consistent with previous findings (Dalgarno, Bishop, Adlong, & Bedgood, 2009).

Conclusions

In science learning, students often experience difficulties in engaging in fruitful inquiry learning, such as generating hypothesis and carrying out scientific reasoning. Despite the availability of various kinds of support or guidance (e.g., prompts), learners may still find it cognitively demanding to successfully complete inquiry tasks. This study proposed and investigated the influences of a 3DCM approach that allowed learners to articulate information about a problem, relevant concepts and their relationships, and the processes of hypothesizing and reasoning about the problem in a holistic picture to support inquiry learning.

The findings show that the students displayed high academic achievement, positive attitudes, low anxiety, and medium levels of confidence, indicating that the proposed approach is a promising means of supporting inquiry learning. Specifically, the 3DCM approach provided the learners with an overview of the inquiry task, and guided them in generating hypotheses step-by-step and developing evidence-based reasoning based on relevant data and knowledge. In addition, the students of high-, medium-, and low-levels showed significant improvements on the knowledge test. Particularly, students of a low academic level benefited the most from this approach, indicating the effectiveness of 3DCM in narrowing the gap between low and high academic level students. Moreover, both the participants’ attitudes and their confidence were positively related to their inquiry skills (including hypothesis generation and reasoning).

Implications

This study has several implications for designing support/guidance for inquiry learning in computer-supported environments. First, it is important to scaffold students when they engage in a complex inquiry task. Visually representing diverse aspects of inquiry learning can facilitate students’ inquiry. Second, narrowing the academic gaps between low and high level students is critically important (Yerrick, 2000; Zohar & Dori, 2003). Providing appropriate inquiry learning environments and external representations, such as cognitive mapping, seems a promising means of realizing the potential of low academic level students and narrowing the gap between low- and high-level students. Third, the impact of the proposed approach on emotions and motivations has implications for the design of learning support for inquiry learning. In particular, reducing learners’ anxiety and improving their confidence is critical to enhance their inquiry skills and knowledge.
Limitations and future work

Some issues or limitations exist in the current study. First, comparison of groups with and without the use of 3DCM is needed to further verify the validity of this intervention. Second, this study investigated only the students’ outcomes on knowledge test, and their perceptions of the inquiry process; it did not examine the quality of the maps constructed, group reports, the discussion process, and the relationship between the quality of the maps and the group reports. Investigation of these components may provide a deeper understanding of the mechanisms of the 3DCM approach. Third, inquiry skills were assessed with a questionnaire rather than a more authentic measure. These issues will need to be addressed in future work.

Acknowledgements

This research is supported by the General Research Fund (No. 17201415) from the Hong Kong SAR Government, and the Seeding Fund for Basic Research (No. 201611159071) from the University of Hong Kong. The EcoMUVE curriculum was supported by the Institute of Education Sciences, U.S. Department of Education, Grant No. R305A080514 to Chris Dede and Tina Grotzer.

References


Leveraging Students’ Knowledge to Adapt Science Curricula to Local Context

Lana M. Minshew*, Kelly J. Barber-Lester, Sharon J. Derry and Janice L. Anderson
Learning Sciences & Psychological Studies, School of Education, The University of North Carolina at Chapel Hill, NC, USA // minshew@live.unc.edu // kelba@live.unc.edu // derry@unc.edu // anderjl@unc.edu
*Corresponding author

ABSTRACT

Conceptions of ecological processes such as the flow of energy and cycling of matter in an ecosystem are increasingly important understandings in a rapidly changing world. This study utilizes a p-prims, or knowledge in pieces, lens to examine understandings and disconnections in students’ conceptualizations of energy flow and matter cycling specific to our context. Findings from our analysis drove continued refinement of our “Compost” curriculum through modifications designed to build on students’ p-prims and foster deeper understanding of ecological processes.

Keywords
Design-based research, Knowledge in pieces, Prior knowledge, Ecological processes

Introduction

Now, more than ever, dilemmas such as pollution and climate change have made it imperative that students become knowledgeable about ecosystems. The Next Generation Science Standards (NGSS Lead State, 2013) align with these growing concerns, having included several ecology standards at the elementary and middle school levels.

In order to fully grasp ecological issues, students need knowledge of complex phenomena such as the cycling of matter and the flow of energy within ecosystems. Research indicates (e.g., Bell-Basca, Grotzer, Donis, & Shaw, 2000; Hogan & Fisherkeller, 1996; Honwad et al., 2010) that ecological concepts such as those require knowledge from both life and physical sciences. Hmelo-Silver and Azevedo (2006) acknowledge the cognitive challenges students face when engaging with complex systems such as ecosystems. Consideration of the vast knowledge required for expert-like understanding of ecosystems as complex systems suggests that student conceptual understanding will necessarily develop over time. Delving deeper into the intricacies of scientific understanding can be a challenge for students due to their limited experiences with these complex phenomena. Yet, as with all learning, student prior-knowledge is important for science understanding. Students’ experiences in local ecosystems contribute to prior conceptions of ecological phenomena that they bring to the classroom.

An emerging approach from the Learning Sciences is diSessa’s (1988, 1993; 2014) Knowledge in Pieces (KiP) framework, which frames student prior-knowledge from a resource-oriented perspective. The KiP framework posits that students have knowledge fragments or pieces of scientific understanding and that restructuring the fragments will aid in development of a deeper, more complete understanding of the scientific phenomena (diSessa, 1988). Student prior understanding is considered primitive, as it is generally an everyday understanding, as opposed to an expert scientific understanding. This perspective counters the historically prevalent misconceptions perspective (Hammer, 1996), which views student non-scientific understanding as incorrect and an obstacle to be overcome.

The KiP framework has been used to understand and analyze student thinking in the context of physics, but has yet to be applied to the study of ecology, specifically the cycling of matter and the flow of energy in an ecosystem. The goals of this study were to examine student conceptualization of these two ecological phenomena through a KiP lens and then determine how student understanding could be leveraged for design modifications to enhance deep learning. Our study examined data collected prior to and directly after the implementation of a curriculum called “Compost.” Compost is a project-based, technology enhanced biology curriculum that focuses on the flow of energy, sustainability, and the cycling of matter in ecosystems. The research questions that guided this study were (1) What KiP related to flow of energy and cycling of matter in ecosystems do students bring with them to the classroom? (2) How can consideration of students’ KiP inform curriculum design, modification and implementation to promote deep learning about energy and matter in ecosystems?
Literature review

Knowledge in pieces

Constructivism posits that students build their understanding of the world around them as they engage in activities both socially and alone. The constructivist approach suggests that learning occurs in phases and recognizes that expert understanding is built up gradually over time (Smith, diSessa, & Roschelle, 1994). From this perspective, student prior knowledge is valuable and initial learning phases are expected to include many incomplete and partial understandings. As existing knowledge is challenged and students are provided opportunities to refine and transform their knowledge, their understanding becomes more sophisticated (Smith et al., 1994). DiSessa developed the KiP framework with these characteristics of constructivism in mind. According to diSessa (2014), students have thousands of primitive ideas or understandings which he calls phenomenological primitives (p-prims). P-prims are loosely organized and highly contextual to student experience (diSessa, 2014).

KiP describes student incorporation of new concepts as a “rewaving” of knowledge (diSessa, 2014, p. 98). In other words, activated p-prims become interconnected with new information and transform student conceptual understanding; p-prims are not eliminated or eradicated. This is in contrast to misconceptions literature that suggests that learning is a process of replacing misconceptions with expert knowledge in a rapid fashion (Smith et al., 1994).

DiSessa and colleagues have done considerable work utilizing KiP in the realm of physics (e.g., diSessa 1988; diSessa, 1993; diSessa, 2014; Smith et al., 1994). These studies consider student intuitive ideas about physics and how those ideas influence formal school instruction of physics. The research recognizes that students interact with the physical world around them and that these experiences influence student understanding and explanation of the physical phenomena.

Knowledge in pieces in life and earth sciences

KiP has been used sparingly for the study of earth and ecological sciences. Two studies directly reference KiP and discuss its merits within the field of ecology. Parnafes (2012) applied KiP to students developing explanations about the earth science concept of the phases of the moon. Parnafes (2012) suggests that because KiP reduces the grain size of analysis concerning student understanding (i.e., p-prims), it highlights the information that needs to be reworked to construct a scientific explanation. Özkan, Tekkaya, and Genan (2004) found that KiP was more appropriate to use than misconceptions when discussing student conceptual understanding of ecological concepts.

Theoretical grounding

The current study is framed by constructivist assumptions that learners’ prior knowledge is an important facet of learning and that learners are active in developing knowledge for themselves (Smith et al., 1994). Constructivism emphasizes the role of prior knowledge in learning and accepts scientific inaccuracies as characteristic in students’ initial phases of learning (Smith et al., 1994). DiSessa’s KiP provides a framework upon which to discuss student conceptual understanding of ecological processes and gradual construction of scientific understanding. It utilizes the constructivist tenet of valuing student prior knowledge and emphasizes the gradual development of student conceptual understanding of complex topics such as ecological processes.

Methods

Research design

This study is a part of a larger National Science Foundation (NSF) supported Design-Based Research (DBR) project called Biosphere. The project aims to develop a middle grades science unit called Compost. Compost is a collaborative, technology-supported, inquiry Biology curricular unit focused on energy and sustainability issues that is suitable for under-resourced middle schools in rural areas of the United States. DBR is characterized by iterative cycles of data collection, analysis, and reflection to inform the design of educational innovations and advancement of educational theory (Easterday, Lewis, & Gerber, 2014).
An important element of DBR is the nature of partnership between researchers and practitioners that drives the design and implementation of the work in authentic context. For this project, the cooperating teacher contributed to the initial design of the curriculum, and has had ongoing involvement throughout development and testing.

DBR as a research approach offers flexibility of implementation and design to fully understand the context in which a study is conducted. Easterday and colleagues’ (2014) model (Figure 1) has six distinct and iterative phases: Focus, Understand, Define, Conceive, Build, and Test. Each phase produces important research and is used iteratively to refine and advance both iteration and theory. Our study reports on the Test phase in which researchers, along with a participating classroom teacher, implement and learn about the developed innovation Compost. The data collected during this Test phase was analyzed to refine the design problem and to gain an understanding of student conceptual understanding concerning the cycling of matter and the flow of energy in an ecosystem.

Figure 1. Steps in design-based research (Easterday, Lewis, & Berber, 2014)

Participants

All participants in the overarching DBR project were in the sixth grade (n = 94) at a small rural science, technology, and engineering focused, one-to-one technology, middle school in the southeastern portion of the United States. Students were from demographically diverse backgrounds: 31% white, 47% African American, 18% Latino/a, 3% Bi-racial, and 1% Asian. The school was identified as a Title I school, indicating that over 40% of the student population qualified for free or reduced lunch. The students all had the same sixth grade science teacher. Of those 94 students participating in the overarching DBR project, 12 focal students were selected by the teacher for deeper study. These 12 students were purposefully selected to reflect diversity of the student population both academically and demographically.

Compost curriculum

The Compost curriculum is an 8-week long, project-based, technology-enhanced inquiry science curriculum. Students build, collect data on, and modify a compost bioreactor in order to develop compost that decomposes quickly. Students utilize computer simulations to run tests on virtual compost piles and collect secondary inquiry information via an online reference tool. The curriculum also consists of several hands-on activities that provide students with experiences surrounding ecosystems, the cycling of matter, and the flow of energy.

Data sources

Model-based assessment

Consistent with our constructivist theoretical framework, we elected to utilize an open ended, Model-Based Assessment (MBA). Using the MBA is consistent with previous research that utilizes model building to demonstrate student conceptual understanding concerning the flow of energy and the cycling of matter in ecosystems (e.g., Hogan, 2000; Lin & Hu, 2003). These studies were able to effectively capture patterns and trends expressed at pre- and post-time points by having students construct models to demonstrate their knowledge concerning the cycling of matter and the flow of energy in ecosystems. The MBA is one way to capture student understanding of complex systems. Complex systems, such as ecosystems, have multiple levels
of organization (Hmelo-Silver & Azevedo, 2006) and it can be difficult to capture student understanding of this kind via a standard assessment.

During the development of the Compost curriculum, the research team created a consensus model of concepts and relationships relevant to energy and matter that were central to the unit (See Figure 2). The consensus model was developed in a three phase process. First, the research team, including two science experts, developed a list of concepts central to the topic areas of flow of energy and cycling of matter that would be expected to fall within the scope of the curricular unit. From there a member of the team drafted the model, which was then presented to the team for review and feedback. In the final step, changes were made to the model based on team feedback. The consensus model was then used to develop the MBA by focusing on creating an assessment that would elicit evidence of student understanding of key concepts and relationships relevant to energy and matter.
The MBA was open and flexible, allowing students to show their understanding in ways reflective of their own knowledge construction. The MBA provides information about deeper understanding by requiring students to apply and represent their own understandings of key concepts without the use of pointed prompts or questions. Figure 3 is a sample of one student’s MBA.

Figure 3. Student example of model based assessment

The MBA was administered by the cooperating teacher during students’ regular science instructional time, and was completed prior to and after the implementation of the Compost curriculum. In addition to being read aloud, the instructions and scenario were on the paper provided. Students were supplied space to generate their model along with a set of stickers, which included elements and organisms common to the scenario ecosystem. Students could use their own drawings and the stickers to depict their model of how energy and matter moved and cycled within the ecosystem.

Interviews

The 12 focal students were interviewed about their MBA within one week of students completing their MBA prior to and after the implementation of the Compost unit. Interviews were conducted to capture student thinking about energy and matter in ecosystems beyond what could be interpreted from examining the MBA in isolation. Students were prompted to explain their MBA to the interviewer. Upon completion of their explanation, students were then probed about specific components (i.e., What do the arrows represent? Can you explain to me about the energy in this model?). Interviews were video recorded to capture the full extent of the interview (i.e., student comments and gestures).

Analysis

The pre- and post-interviews of 11 students were transcribed and analyzed, totaling 22 interviews in all. The 12th student was unavailable for a post-interview due to numerous school absences. That student’s pre-interview was, therefore, dropped from the analysis due to incomplete data. The data were analyzed through interaction analysis, coding and stepwise reduction through data display and summary analysis.

Interaction analysis

The initial phase of analysis included Interaction Analysis (IA; Jordan & Henderson, 1995) for three of the student interviews. Interaction analysis is an in depth process that involves repeated viewing and noting of video
selections by a team of researchers. Audio-recordings of the IA sessions and session notes were analyzed to identify salient themes within the students’ interviews. Five major themes emerged from the analysis of the IA sessions: energy, matter, food chains, decomposers, and waste.

Coding

In the second phase of analysis the entire corpus of twenty-two interviews was coded for student p-prims using the themes identified in the IA sessions as initial broad start-codes. Interview segments were coded with the unit of analysis for each code typically being the talk-turn, but sometimes being a portion of a talk turn when multiple concepts were discussed within a single talk-turn. For the first step of the coding process, two members of the research team coded a single interview. Following the independent coding, the researchers cross-checked their application of the start codes. The cross-check process confirmed that both researchers applied the codes consistently, with no disagreements. A single researcher coded the remaining twenty-one interviews. Table 1 depicts the five major topic areas and the associated p-prims along with student quotes to illustrate the p-prims.

<table>
<thead>
<tr>
<th>Topic area</th>
<th>P-prims</th>
<th>Student quote from pre-interview</th>
<th>Student quote from post-interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food chains</td>
<td>There is a relationship between sun/sunlight and producers</td>
<td>“Okay, so I connected the plants to the light, because the plant is a producer.” (Mireya)</td>
<td>“I put the producers and sun and water together because that is what the producers need to grow.” (Mireya)</td>
</tr>
<tr>
<td></td>
<td>Animals eat other animals.</td>
<td>“I made a food chain to show how other things eat off other things to get food and energy. So I made a plant, then a rabbit eats the plant, then a snake eats a rabbit, then a fox eats the snake.” (Jaylen)</td>
<td>“Well the energy flows from the grass to the rabbit, from the rabbit to the snake, from the snake to the owl.” (Olivia)</td>
</tr>
<tr>
<td></td>
<td>Animals eat plants.</td>
<td>“The deer needs the producers so they can eat.” (Francisco)</td>
<td>“Producers make food for themselves and other living things.” (Jaylen)</td>
</tr>
<tr>
<td></td>
<td>Decomposers are a part of the food chain.</td>
<td>“Then the fox eats the snake, and humans eat other animals, but bacteria breaks down other foods and things so that’s why I put that at the top.” (Jaylen)</td>
<td>“It starts from the producer, then it goes to the consumer, then decomposers. The producers make the food and then the consumers eat the producers then the decomposers eat the consumers and the plants.” (Mireya)</td>
</tr>
<tr>
<td>Decomposers</td>
<td>Bacteria, worms and mushrooms are decomposers.</td>
<td>“The mushroom is a decomposer, the worm is a decomposer too.” (Malcolm)</td>
<td>“So like when you have waste and they [mushrooms] can decompose it. Then it goes back into the soil and gives the soil all rich nutrients” (Malcolm)</td>
</tr>
<tr>
<td></td>
<td>Decomposers break things down.</td>
<td>“Mushrooms and bacteria um break down waste” (Francisco)</td>
<td>“You will need the worm to break down the waste and turn it into soil and the soil can be used to create plants.” (Tori)</td>
</tr>
<tr>
<td></td>
<td>Decomposition is connected to compost.</td>
<td>N/A</td>
<td>“So I tried my best to draw a compost pile. I drew included all like the greens and the browns 1, 2, 3. Then I drew several decomposers around it.” (Cierra)</td>
</tr>
</tbody>
</table>
Waste

Animals and humans generate waste.

“The deer would leave the waste and like the waste would help the soil.” (Olivia)

Some waste can be broken down and is connected to soil.

“Bacteria break down waste so that plants can use it.” (Francisco)

Energy

The sun is central to energy.

“It is basically all about the sun, the energy and how it spreads out for the whole forest.” (Leandra)

Rain and clouds are connected to energy.

“Rain is the energy for the plant because it makes the plants grow.” (Jaylen)

Organisms need energy.

“The sun umm helps us get energy and it goes to the plants.” (Fatima)

Electricity is connected to energy.

“So I did that to show that the energy, the battery connected to the light pole will be energy because the battery helps the light come on.” (Devon)

Matter

Everything is matter

“I mean all of it is matter because everything is made up of matter, but then I don’t know.” (Olivia)

Matter exists in different states.

“Liquid … [points to water sticker] solid [points to the human sticker] … gas [points to oxygen sticker].” N/A

Results

In response to our first research question, our analysis demonstrated that KiP is a useful and applicable framework for understanding and analyzing student knowledge related to ecosystems. Analysis of the interviews revealed six main areas in which students held various p-prims. The following sections present student p-prims within each topic area from both pre- and post-interviews highlighting shifts in these p-prims evidenced in the student interviews.

Food chains

The topic of Food Chains was prevalent within student interviews; eight of the eleven students spoke at length about the topic at both time points. Student explanations related to food chains included both implicit and explicit references. A few students explicitly used the term “food chain,” while others included descriptions of elements of food chains (e.g., producer-consumer relationships) without explicitly naming their depiction as a
food chain. Changes from pre- to post-interviews centered on student ability to provide details about food chains and the frequency with which they spoke about food chains. The following are p-prims related to food chains found during analysis.

There is a relationship between sun/sunlight and producers

Six students connected the sun or sunlight to plants or producers during their pre-interviews. The nature of the relationship described between the sun and producers varied among students. Some students simply connected the sun to plants without specifying the nature of the relationship, while others articulated the connection between the sun and plants in terms of the energy that the sun provides to plants. For example, Mireya stated, “Okay, so I connected the plants to the light, because the plant is a producer.” However, not all students expressed this p-prim in their pre-interviews. In contrast, during their post-interviews students’ connection between the sun and plants was slightly stronger with eight of the eleven students making the connection. In her post-interview Mireya said, “I put the producers and sun and water together because that is what the producers need to grow.” She was able to provide a few more details about why the producers and the sun should go together. Other students held p-prims that connected the sun to plants in their post-interviews whereas they had not made this connection in their pre-interviews.

Animals eat other animals

In their pre-interview explanations, eight students put a major emphasis on the chain of consumption of smaller animals by larger animals. Common examples included birds eating worms and snakes eating rabbits. The idea of consumption was connected by two students to animals’ need to consume other animals to obtain energy. Jaylen noted, “I made a food chain to show how other things eat off other things to get food and energy. So I made a plant, then a rabbit eats the plant, then a snake eats a rabbit, then a fox eats the snake.” Jaylen indicated that the food chain not only showed animals consuming other organisms for food but that this was how energy was obtained. Whereas Mireya stated, “I connected the owl and the snake together because I tried to make it look like a food chain,” demonstrating that this student had a p-prim for animals eat other animals but it was not as developed as Jaylen’s p-prims. Animals eating other animals continued to be a prevalent idea expressed in student post-interviews. Olivia was able to articulate, “Well the energy flows from the grass to the rabbit, from the rabbit to the snake, from the snake to the owl” demonstrating that she knew that when animals consumed other organisms they obtained energy which she did not discuss during her pre-interview.

Animals eat plants

Similarly, the same eight students provided examples of animals eating plants. The most common example included was rabbits eating grass or plants. The consumption of plants by animals was connected by several students to the animals’ need for energy, which the plants provided. Several students used the words “producers” and “consumers” to describe the relationship between plants and animals. Francisco demonstrated this p-prim when he said, “the deer needs the producers so they can eat.” This p-prim remained relatively constant from pre-to post-time points, with one addition being that students articulated that plants produced their own energy and provided the energy for consumers during their post-interviews. For example, Jaylen noted in his post-interview, “producers make food for themselves and other living things.”

Decomposers are a part of food chains

Jaylen was the only student to connect decomposers to food chains during the pre-interview. However, five students were able to connect decomposers to food chains during their post-interview. Mireya, for example, stated in her post-interview, “The producers make the food and then the consumers eat the producers then the decomposers eat the consumers and the plants.” Mireya was able to connect producers, consumers, and decomposers, together to construct her conception of a food chain, which in her pre-interview only consisted of stating, “an owl and a snake are connected.”
Decomposers

Decomposers and the process of decomposition is central to the Compost curriculum. When the curriculum was developed it was anticipated that students would have nominal understanding about decomposers and decomposition. The interview and analysis process revealed that eight of the eleven students acknowledged decomposers in their MBA pre-interview and all eleven students discussed decomposers in their post-interview. The following is a discussion of the p-prims associated with decomposers and decomposition.

Bacteria, worms, and mushrooms are decomposers

At the time of the pre-interview three students had p-prims that allowed them to correctly identify bacteria, worms, and mushrooms as decomposers. Other students named bacteria, worms, and mushrooms as being a part of the ecosystem depicted in their MBA’s but did not identify these organisms as decomposers. For example, Fatima commented, “bacteria is mostly kind of everywhere on the ground.” Fatima has a p-prim that puts bacteria into the environment but she does not articulate bacteria’s role within the environment. In their post-interviews nine students directly identified bacteria, worms, and mushrooms as decomposers while the remaining two students used the general term decomposer. Tori did not talk about decomposers at all in her pre-interview nor did she have them depicted in her pre-MBA. In her post-interview, however, Tori specifically included and identified bacteria, worms and centipedes as decomposers.

Decomposers break things down

Five students had p-prims at the pre-interview associated with the idea that decomposers break things down. Their expressed understanding ranged from stating simply that decomposers break things down to explaining that bacteria break down waste and put organic matter back into the soil. Francisco expressed, “mushrooms and bacteria um break down waste.” Francisco does not identify the organisms as decomposers but has the p-prim that the organisms serve the purpose of breaking down waste. In contrast, Cierra demonstrated stronger organization of p-prims articulating that, “the bacteria and mushrooms … they break down, if an animal dies or something like that it will decompose the dead body into like organic matter that the soil to use and then it makes the ecosystem even healthier.” Nine students had the p-prim that decomposers break things down at the post-interview. With the majority of students indicating that the broken down material is put back into the soil to improve the environment.

Decomposition is connected to compost

Three students introduced a third p-prim, connecting decomposition and compost, in their post interviews. Jaylen identified waste, organic matter, greens, and bacteria as elements within compost. Cierra drew a compost pile in her MBA with several decomposers surrounding it. Tori referenced the connection between decomposition and the compost bottles that students created and studied during the Compost curriculum.

Waste

The majority of the students interviewed incorporated waste into their models and explanations in some form in both their pre- and post-MBAs, yet how they discussed the concept of waste differed. Of the students that included waste in their explanations, incorporation ranged from only including a mention of trash to a fleshed out comparison between organic and inorganic waste. Two p-prims related to waste were identified.

Animals and humans generate waste

Waste generated by humans was most frequently connected to the idea of “trash.” Students used the word “trash” specifically on multiple occasions, and provided examples of things that were human generated waste, such as a plastic bag. Students connected animal waste to the idea of feces. One of the stickers provided for students to choose from was a picture of feces labeled “waste.” Students frequently connected animals to this sticker, and used the word “waste” to describe the sticker. This p-prim was salient and changed little from pre- to post-interviews with eight students discussing the p-prim at both time points.
Some waste can be broken down and is connected to soil

Three students in their pre-interviews connected animal waste with decomposition. As demonstrated earlier, students noted that decomposers break down animal waste. Francisco articulated that, “bacteria break down waste so that plants can use it,” while Olivia noted that animal waste “would help the soil.” The process of decomposition tended to be connected to animal waste and not to human waste. One student, Cierra, exhibited a much more advanced explanation during her pre-interview, differentiating between “organic matter” and “inorganic matter.” Cierra explained, “organic is the stuff being able to be decomposed and inorganic as potentially harmful to the environment, “because the decomposers cannot decompose it.” The number of students expressing the p-prim that waste can be broken down and is connected to the soil increased from pre- to post, with eight of the eleven students articulating the connection. Cierra maintained her p-prim for the difference between inorganic and organic matter while other students demonstrated primitive conceptions of inorganic and organic waste. For example, Mireya stated, “We produce this waste [points to the sticker displaying feces] and this is like when people just throw trash out [points to the sticker depicting a trash can].” Mireya did not use the terms organic and inorganic when describing the scenario, but her statement does indicate that she recognizes that there is a difference between the two types of waste.

Energy

The flow of energy in an ecosystem is a complex topic. Food chains, as mentioned above, are salient in students’ understanding of how energy is passed from organism to organism in an ecosystem. In addition to the p-prim related to food chains, evidence for four distinct p-prim related to energy was present in student interviews.

The sun is central to energy

Six students expressed this idea during their pre-interviews. This idea was loosely connected to the p-prim that plants need the sun in order to grow. The sun as central to energy was more salient in student post-interviews with 10 of the 11 students expressing the p-prim. Student complexity of understanding ranged during their post-interviews; some students identified that the sun is energy that is needed in an ecosystem whereas others connected the sun to plants and the process of photosynthesis.

Organisms need energy

Six of the eleven students articulated their understanding that organisms need energy during their pre-interview. These p-prim tended to be simple. For example, Fatima stated, “the sun helps us get energy and it goes to the plants.” Fatima noted that humans and plants need energy, but did not specify why the energy is needed. This p-prim progressed and during student post-interviews more students, ten of the eleven, mentioned that organisms need energy, and five of those students expressed that energy flows within the ecosystem. Olivia stated in her post-interview, “Well, the energy flows from the grass to the rabbit, to the rabbit to the snake, and from the snake to the owl.” Demonstrating that this p-prim became more organized and demonstrated a deeper understanding of knowledge concerning energy and organisms.

Rain and clouds are connected to energy

Though they were in the minority, two students related rain and/or clouds to energy during their pre-interviews. Jaylen explained that “rain is the energy for the plant because it makes the plants grow.” He continues to explain that “oxygen, heat, sunlight, and rain give flowers the energy,” indicating that perhaps an underlying connection to photosynthesis may support this connection. During his post-interview Jaylen incorporates additional information, the inclusion of carbon dioxide, with his p-prim about rain and clouds. He states, “sun and water drops help plants grow, and oxygen and carbon dioxide in the air also helps.” Jaylen’s inclusion of carbon dioxide is one step closer to the process of photosynthesis. Ethan, however, was less specific about the connection between rain, rain clouds and energy during his pre-interview. When asked about what he might add to his pre-MBA, Ethan responded, “make like rain clouds for more energy… .”
Electricity is connected to energy

While nearly all students related their discussion of energy exclusively to elements traditionally associated with ecosystems, such as the sun, plants and animals, Devon brings up batteries, lights and light poles as related to electricity during his pre-interview. The student does not explicitly use the word electricity, but the elements discussed all related to electricity. In Devon’s post-interview he no longer discusses the idea of batteries but includes the p-prim that the sun’s rays hit solar panels and the panels can provide light. Devon stated, “Okay, first I drew the solar panels, the sun helps the solar panels and gives energy to the solar panel, the solar panel gives energy to light.” This student was the only one to connect the concept of energy to things related to electricity, but the connection points to an area of experiential knowledge from which many students would likely be able to draw.

Matter

The cycling of matter in an ecosystem is another central theme in the Compost curriculum. When asked specifically about matter in their models, six students during their pre-interview either did not directly respond to the question or simply stated, “No.” Jaylen responded similarly when questioned about matter, stating, “I don’t think I really think about the matter.” The number of students who did not speak directly about matter decreased to three at the post-interview.

Everything is matter

Students who responded to the question about matter in the MBA did so with uncertainty at the pre-interview. Several students activated the p-prim that everything is made up of matter, but verbal hedging around statements about matter revealed their tentativeness. For example, Olivia stated, “I mean all of it [items included on her MBA] is matter because everything is made up of matter, but then I don’t know.” Ethan’s response was also shrouded in vagueness. Ethan responded by stating, “Matter is, it’s like it takes up like it takes up a big part of this [the MBA] because it is a lot of space taken up in there.” Student uncertainty diminished during post-interviews, however, their p-prims did not advance in complexity. Olivia still confirmed “matter makes up everything so matter is everywhere.” She, however, expressed this idea with less hesitation than in her pre-interview.

Matter exists in different states

Devon responded to the prompt about matter during his pre-interview by identifying the different states of matter represented in his MBA. Devon’s interpretation of the prompt was unique, in that none of the other students identified the states of matter. This does not mean that the other students did not possess the p-prims related to states of matter, but rather that they did not activate this specific p-prim when they were constructing their pre-MBA.

Discussion

The analysis presented here responds to our first research question by outlining how KiP proved to be a useful and applicable framework for understanding and analyzing student knowledge related to ecosystems. Analysis of the pre-interviews and models showed that students bring with them p-prims in distinct topic areas. These p-prims can be leveraged in science classrooms to support student deeper understanding of the content material. Analysis of the post-interviews allowed us to see how student p-prims shifted after having participated in the Compost unit. In response to our second driving question, we now address how student p-prims can be used in order to support curricular design modifications to promote deep learning about the flow of energy and cycling of matter in ecosystems.

Language emphasis on matter

Our analysis revealed that students had an array of p-prims about the flow of energy and the cycling of matter in ecosystems. For example, some students activated p-prims that expressed that matter is everywhere, or that
everything is made up of matter consistently from pre- to post-interviews, but did not discuss the process of cycling of matter. Many students also possessed p-prims about decomposers breaking down material and putting it back into the soil. Student knowledge was fragmented, preventing them from connecting the topic of matter to the topic of decomposers breaking down waste. One curriculum-design modification for deeper learning is to include more targeted language in the curriculum to support students in making this connection. The purposeful inclusion of phrases such as “cycling of matter,” “rate of decomposition,” and “process of decomposition” can be used to focus student attention and engage students in the discussions about matter and the cycling of matter. These phrases can help to make the topic of matter more salient and explicit in student thinking and provide opportunities to incorporate new information with existing p-prims.

Connecting matter to food chains

Analysis of student p-prims revealed that the Compost curriculum did not engage students in the topic of food chains beyond their existing p-prims. As indicated above, students expressed the understanding that organisms are connected to one another in an ecosystem through a mechanism called a food chain. Students focused on the linear nature of the food chain and on organism to organism food consumption with several students connecting energy flow with food consumption. The nature of students’ food chain p-prims did not reflect much connection to the cycling of matter in an ecosystem. A design modification which could help students activate, reorganize and extend their p-prims related to matter and food chains would be to incorporate a purposeful emphasis on the composition of organisms and carbon and nitrogen cycling. Activities would center on answering question like: What are we made of? Where does the stuff we are made of come from? Those same questions could be extended to other organisms as well, drawing students into a deeper understanding of the complex relationships between energy flow and matter cycling.

A deeper look at photosynthesis

Students had p-prims that connected the sun to energy and the sun or light to plants. However, the topic of photosynthesis was only introduced formally by one student, Jaylen, and was merely hinted at by others. Photosynthesis is an important topic when discussing the flow of energy in an ecosystem, and also provides an opportunity for a micro-level look at how matter moves in an ecosystem. A new activity could be developed and incorporated into the Compost curriculum that engages students with the topic of photosynthesis diving deeper into the concept of energy flow in an ecosystem. For example, a collaborative photosynthesis model-building activity could push students to follow the atoms and molecules before and beyond the moment of photosynthesis to encourage students to consider the interconnections between energy flow and matter cycling.

Teacher support material

Finally, the identification of student p-prims from both pre- and post-interviews will be important in the development of the teacher support material. By having advanced knowledge of the types of p-prims students bring with them, teachers will be able to aid in students’ “re-weaving” (diSessa, 2014). As designers, we can incorporate questions, prompts, and hints into our teacher e-book to provide guidance for teachers when they encounter student p-prims. For example, as evidenced by their interviews students were able to identify different types of decomposers; bacteria, mushrooms, and worms. These decomposers can be identified as either micro or macro. By providing teachers with information on the differences between these decomposers that students can readily identify, teachers can delve deeper with their students on the roles decomposers play in ecosystems and the impact different decomposers have on the process of decomposition. This type of guidance may help students to organize and re-weave their p-prims.

Conclusion

We have suggested that Knowledge in Pieces (KiP) is a suitable way to discuss students’ emerging scientific understanding in the domain of ecology. KiP states that students have thousands of loosely organized, highly contextual, and fluid p-prims that shape how students understand scientific concepts. The loose organization of p-prims allows student understanding to be reshaped through phases as new information is gathered. KiP stands in direct contrast to a misconceptions perspective, which views student prior understanding as deeply rooted,
resistant to change, and as something that must be overcome in order for scientific understanding to occur (Hammer, 1996; Smith et al., 1994).

Our findings suggest that identifying student p-prims is valuable for designers in order to generate curriculum that encourages deeper understanding of scientific concepts. Student p-prims are not a “road block” to student learning, but a valuable insight into student conceptual understanding (diSessa, 2014). Leveraging student ideas and incorporating them into curriculum encourages students to push their primitive understanding closer to that of experts.

Though the small sample and unique rural context limit the generalization of the specific p-prims uncovered, the successful application of the KiP framework to the ecological topics of flow of energy and cycling of matter is an important addition to the field. The goal of this study was to examine student initial conceptual knowledge related to ecological processes with the KiP framework. The data provided by the interviews offered an in-depth look at the p-prims related to ecological processes that our students brought with them to the classroom. Our design proposals exhibit how that knowledge of student p-prims could be leveraged in curricular design to support deeper understanding. The pursuit of demanding, inquiry-based science instruction for fostering students’ deeper understanding of complex science topics should be a major priority for all of us who seek to prepare our students for the challenges of today and the future. The KiP framework provides an important lens for building our own understanding of students’ thinking in order to more productively design for and facilitate their knowledge building processes.

Acknowledgements

We thank our participating teacher and students. This material is based upon work supported by the National Science Foundation under Grant No. DRL1418044. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References


Moving Apart and Coming Together: Discourse, Engagement, and Deep Learning

Andrea S. Gomoll1*, Cindy E. Hmelo-Silver1, Erin Tolar1, Selma Šabanović2 and Matthew Francisco2

School of Education, Indiana University, Bloomington, IN, USA // 2School of Informatics and Computing, Indiana University, Bloomington, IN, USA // agomoll@indiana.edu // chmelosi@indiana.edu // etolar@iu.edu // selmas@indiana.edu // francm@indiana.edu

Corresponding author

ABSTRACT

An important part of “doing” science is engaging in collaborative science practices. To better understand how to support these practices, we need to consider how students collaboratively construct and represent shared understanding in complex, problem-oriented, and authentic learning environments. This research presents a case study centered on the work of four students in a human-centered robotics curriculum enactment. We explore how discursive features including embodied gesture and positioning of material artifacts contributed to the problem-solving process and helped students move towards deeper learning — showing how nonverbal and verbal discourses were used to construct agreement and disagreement, parallel interaction, and accountability. Each of these discursive actions informed how the group moved forward or was halted in their complex collaborative work. We found that early stages of constructing a joint problem-solving space (JPSS) in this classroom environment required extended engagement, student ownership, and negotiation of shared activity. By exploring how select students worked toward the co-construction of joint problem-solving spaces, we reimagine what deep engagement and learning in STEM learning environments can look like, and we inform better design for the creation of these spaces.

Keywords

Human-centered robotics, Joint problem solving, Collaborative learning, Discourse analysis

Introduction

As learners construct knowledge in social and interactive contexts, they negotiate power dynamics and disagreement, establish group norms, and navigate what it means to come to a consensus (Barron, 2000). In this article, we focus on these complex collaborative learning practices as a central piece of deep STEM engagement and learning. Collaborative learning practices are essential to STEM interest, engagement, and learning. These collaborative practices are also important for students’ ability to see themselves as capable of “doing” science and engineering (Tan et al., 2013). Collaboration is essential for STEM careers, which require practitioners to navigate jointly what cannot be solved alone, obtain and evaluate evidence, design solutions, and communicate pertinent information. In the Next Generation Science Standards (NGSS), Engineering Practices 7 and 8 pertain to the ability to obtain and evaluate evidence, construct explanations, and design solutions (NGSS Lead States, 2013). Engaging students in the work of collaborative learning practices is therefore intertwined with authentic STEM engagement in classroom contexts and beyond. We conjecture that supporting students as they engage in these collaborative practices promotes deep learning. Although there is no single definition that we have found to categorize “deep learning,” we view it as learning that leads to robust understanding of disciplinary practices, disciplinary concepts, and how these are interconnected (Sawyer, 2008). To better understand how to support deep collaborative learning, we need to consider how students navigate the challenging task of constructing shared knowledge (Barron, 2000; Roschelle, 1992).

The research presented here reveals the process of constructing joint problem-solving spaces (JPSS) in the discourse of one small group of middle school students. JPSS are shared “conceptual spaces” where students are able to develop their understanding together (Teasley & Roschelle, 1993, p. 69). These spaces, where students negotiate using shared language and activity, can be an important springboard for long-term STEM engagement and deep learning. As students collaboratively construct knowledge, they use verbal and nonverbal discourse to create a shared language for problem solving. This negotiation spurs students to engage deeply with disciplinary content and practices (Engle & Conant, 2002). The interactional data presented here comes from a larger research project that aims to engage students with engineering design practices through the use of a socially-oriented robotics curriculum. Prior studies of JPSS have taken place predominantly in lab settings or with advanced student populations (Hmelo, Nagarajan, & Day, 2000; Roschelle, 1992). Studying JPSS in a complex classroom setting advances our understanding of how students jointly progress toward deeper learning in naturalistic environments.
The human-centered robotics (HCR) curriculum described here is an inquiry-based curriculum of robotics experiences centered on a human-centered problem (“How can we create a robot that serves a need in our local context and connects students with remote peers?”). We designed and studied an HCR curriculum that helps learners develop technical skills along with an understanding of the relationship between technology, nature, and society. A key goal of this curriculum was to engage non-dominant populations (e.g., females and ethnic and racial minority populations) in STEM via a social robotics experience. Hands-on robotics and participatory design activities are an effective way to engage diverse groups of students in STEM (DiSalvo et al., 2008). HCR attends to “societal context” and “human needs” and has led to increased motivation in STEM, particularly for underrepresented female populations (Pajares, 2005). In this robotics unit, students in Indiana collaboratively personalized and programmed robots to both meet a local need and to communicate telepresently (i.e., via remotely controlled robot) with students from Alaska (Gomoll, Hmelo-Silver, Šabanović & Francisco, 2016). This connection was framed as an opportunity to interact with students and physical environments they might not be able to visit otherwise.

As students worked together to build a telepresence robot, they negotiated how to organize the problem-solving process. This negotiation within a real-world context involves active and creative problem solving, and is thus an ideal context for deep learning (Wang, Kirschner, & Bridges, 2016). In this study, deep learning occurs as students move beyond the ability to connect basic robotic components to recognizing how they work together as a designed socio-technical system. It also means that students come to appreciate the potential of collaborative roles and iterative design within the robotics project. Thus, our present analysis aims to understand how and where students make these connections, as well as when they fail to do so. Joint problem-solving spaces (JPSS) are constructed through interaction as group members set goals, define and redefine problems, and propose actions to solve them (Barron, 2000). This negotiation is akin to the kind of problem solving process that STEM practitioners and designers navigate (Silva, 2008). By pinpointing where JPSS are and are not occurring, as well as how the success and failure of JPSS are mapped to discourse, we can better understand how this curriculum supported deep engagement on the way to deep learning. In doing so, we also inform the literature on how to design curricula that support deep STEM learning experiences. This research expands upon earlier work around JPSS as well as research that has been conducted on knowledge convergence — exploring how these collaborative spaces are constructed in the face-to-face learning environment of an HCR experience over an extended period of time, and how they act as a precursor to collaboratively constructed understanding (Jeong & Chi, 2007; Roschelle, 1992). We focus in particular on the nonverbal and embodied contributions and collaborative practices that shaped one small group’s work.

This paper addresses three research questions:

- How do individual students engage with collaborative work within this curriculum?
- How and where do we see JPSS in the collaborative work of one group in a middle school robotics curriculum?
- What discursive practices support and impede the creation of JPSS in this human-centered robotics context?

In the sections that follow, we provide an overview of joint problem solving and engagement, describe our qualitative analysis methods, and present episodes of small group interaction to reveal the dynamic nature of problem solving in one small group’s experience of our HCR curriculum.

**Joint problem-solving spaces and discursive practices**

In early stages of collaboration, learners have different knowledge about the concepts they are working to master, challenging them to construct shared conceptual understanding through creation of a joint problem-solving space. When this joint construction of a space for problem solving is achieved, learners are able to move towards knowledge convergence (Roschelle, 1992). When knowledge convergence occurs, increasingly similar representations and socially shared meaning are produced by group members (Jeong & Chi, 2007; Weinberger, Stegmann, & Fischer, 2007). Jeong and Chi (2007) found that collaboration experiences led to knowledge convergence. However, these authors highlighted that examining conversation alone was not enough to paint a full picture of knowledge convergence as a collaborative process. We must also attend to social norms, nuanced contributions, and indicators of divergence. Before knowledge convergence can occur, students must create a JPSS where new understanding can be built. Within a JPSS, students establish common ground and coordinate their interactions to achieve a shared goal (Barron, 2000). We aim to better understand the work of constructing a JPSS by exploring the interaction of one small group.
As students work together to build a telepresence robot, they make decisions about how to organize the problem-solving process. Working cooperatively, students often take up individual roles and piece their work together to create a final product (Damon & Phelps, 1989). This group structure makes it difficult to construct a shared problem space, as one group member’s ideas may dominate the problem-solving process. Alternatively, groups working collaboratively tackle problems together, offer peer feedback, and build on each other’s ideas (Jordan & McDaniel, 2014). JPSS, and the interactions that construct them, serve as indicators of effective collaboration. If participants do not develop a shared problem space, they may not converge upon a shared understanding, and thus deep learning for all of the participants is unlikely. As demonstrated in Barron’s work (2000), middle school groups engaged in complex problem solving often struggle to agree upon their goals, negotiate joint solutions, and achieve shared understanding. The barriers to engagement in complex problem solving and coordination included focusing on individual outcomes, treating personal workbooks as territory (i.e., only recording your own ideas), and failing to acknowledge peer contributions. High coordination in Barron’s (2000) contrasting case study was marked by consistent turn-taking, shared workbook navigation, productive conflicts, and uptake of peer ideas. These activities all constitute discursive practices — doing something to shape an interactional event in situ (Potter & Wetherell, 1987; Potter & Hepburn, 2008). Barron and Roschelle (2009) argued that instances of disagreement and repair are necessary for the achievement of shared cognition. As groups uncover discrepancies, they must repair them in order to achieve shared understanding. This repair acts as a catalyst for knowledge convergence.

As we work to design a curriculum that supports knowledge convergence and deep STEM engagement, it is important to create spaces where discrepancies are expected and encouraged, and where organic JPSS can emerge. Our work unpacks one group’s patterns of interaction — considering how verbal and nonverbal discourses are connected to a group’s engagement and their eventual trajectory toward the construction of a JPSS and shared understanding.

**JPSS, engagement, and deep learning**

In past research, behavioral engagement (e.g., on-task behavior and participation), cognitive engagement (e.g., willingness to persevere to solve a problem), and emotional engagement (e.g., student attitudes and interests) have been privileged as evidence of students’ interest and learning (Fredricks, Blumenfeld, & Paris, 2004). These forms of engagement are often studied in isolation, when they should be considered as dimensions of the complex and contextually determined concept of engagement (Fredricks et al., 2004). Furthermore, these defined forms of engagement are difficult to measure and interpret in meaningful ways (Fredricks & McColskey, 2012).

Throughout this paper, we consider engagement as context dependent and continually in flux. Across diverse definitions of engagement, there is agreement that it is a multidimensional construct. Engagement is nuanced and involves emotional, behavioral, and cognitive dimensions — including multifaceted forms of participation (Appleton, Christenson, & Furlong, 2008; Fredericks et al., 2004; Sinha et al., 2015). As such, this research examines how engagement unfolds in an HCR learning environment over time and across groups.

Exploring engagement and its connection to the construction of JPSS, we look for evidence of co-construction of understanding as students participate in diverse ways. We attend to the ways that students use verbal and nonverbal discourse to structure their interaction — including the use of gesture and the orientation to physical materials at the group’s table. Such materials can be considered “boundary objects,” artifacts that make the negotiation of group norms and understanding possible (Engeström, Engeström, & Kärkkäinen, 1995). Nonverbal discourse and embodied gesture are important aspects of communication and cognition (Alibali & Nathan, 2012). Human interaction depends on the use of a large toolkit of semiotic resources — from the sequence of talk to the positioning of the body, to aspects of the surrounding environment made relevant through gesture (e.g., pointing; Goodwin, 2000). Researchers have increasingly attended to embodiment (e.g., pointing, gaze, posture) as a significant interactional feature that can be used to perform discursive action such as collaborative knowledge construction and participation (Nevile, 2015). We conjecture JPSS and the nuanced verbal and nonverbal engagement that occurs within them not only to be an indicator of effective collaboration, but also as a precursor to deep learning.
Methods

Participants

Participants in this study were students in a five-week HCR unit in a seventh and eighth grade Applied Science class in a rural, U.S. public school in Indiana. Of the 30 students in the class, the instructor selected eight as focus group participants. One focus group of four was selected for analysis because of more appropriate and respectful group dynamics that the research team viewed as a condition for examining the deep engagement needed for constructing a JPSS. This group included two female eighth grade students and two male seventh grade students. All students were Caucasian. The rural demographic and female population of this middle school class was categorized as a non-dominant population in STEM (Avery, 2013; Tan et al., 2013).

Instructional context

Over the course of the unit, students collaboratively personalized and programmed robots to meet both a need in their local environment and to engage telepresently with students from Alaska (see Gomoll et al., 2016). Telepresence connections allowed students to drive a robot remotely. For example, a student in Indiana would drive a robot in Alaska — engaging with remote peers using a video feed and sensors embedded in the telepresence robot’s design. Students worked collaboratively in groups of four with support from a facilitator as they tackled ill-structured problems. Students also identified a need in their classroom that could be solved by a robot, and adapted their robot to solve this problem. This dual focus between solving a local problem and communicating with distant peers helped students consider how robots could both affect their local environment and be used to connect to a remote context.

![Figure 1. Overview of HCR unit trajectory](image)

Figure 1 shows an overview of the HCR unit. During the unit introduction, students experienced several activities that introduced HCR and basic engineering practices. In groups of four, they moved between tables where different HCR interactive experiences were set (Figure 1A). The design challenge was then presented: Design a robot that serves a local need in your community and allows remote students to explore your space telepresently. Students brainstormed ways that robots could be used to address local needs (1B). Students then created maps of their classroom as aids for telepresent exploration and to better understand how a robot would move in their local environment (1C). This was followed by an embodied programming activity where students discovered the need for clear, specific instructions (1D). Here, they took on the role of the robot and translated...
directions into basic code. Students then practiced programming and troubleshooting as they drove the iRobot Create (1E). In the design portion of the unit, students designed prototypes of robots for their schools (1F). Each group agreed upon a unified design and was challenged to build and program a working human-centered robot, building up from an iRobot Create platform (1G). Throughout the unit, an engineering design cycle (see Figure 2) was used as a resource to help students to organize and reflect upon their problem-solving process. Students were encouraged to reflect on which part(s) of this cycle they were enacting as they designed and built their robots. In final presentations of their robot designs, groups used the engineering design cycle as a framework for explaining their group’s process. Students provided examples of questions asked, information collected, challenges faced in the process of testing solutions, and refinements made based on these challenges.

Figure 2. Engineering design cycle adapted from Resnick (2007)

Data sources and selection

Video of the focus group was recorded for every class session, with data from four class periods selected for this analysis. The selected videos involved group reasoning and joint activity directly related to our research question about the nature of collaborative work in this setting. In these clips, the students were engaged in tasks that could not be completed alone, including embodied programming, troubleshooting, and user testing. We were interested in the construction of JPSS in collaborative tasks, and we therefore selected inherently collaborative tasks to provide the best chance of “seeing” a JPSS emerge.

The first author developed initial analytic memos for each video. These memos noted the actions performed in small group discourse (e.g., disagreement) and features that related to the group’s problem solving. Seven 2- to 7-minute video clips were selected for further exploration. These clips included group problem-solving experiences where a variety of verbal and nonverbal discourse features were highlighted in memos. These moments were flagged as potential JPSS.

Analysis

We investigated our research questions using tenets of discourse analysis. Discourse analysis considers talk as a social action, and can be used to identify discursive patterns and norms that both shape and are shaped by participants (Gee & Green, 1998; Potter & Wetherell, 1987). We used discourse analysis to consider how this group’s nuanced engagement shaped their collaborative inquiry and knowledge construction. We were interested in (a) the role of both nonverbal and verbal discourse in an interaction, and (b) how norms and practices such as turn-taking, agreement, disagreement, and decision-making, evolved at individual and group levels (Edwards & Potter, 1992).
The seven case-study clips were transcribed and analyzed in four data sessions. Participants in the data sessions noted interesting features of each interaction and converged around three themes related to the creation of JPSS: discrepancy, accountability and ownership, and parallel play. These themes were adapted into emergent codes and further memos were then applied to transcripts. Early codes were tags that noted features and functions of talk related to collaborative inquiry (e.g., disagreement, accountability, facilitation, questioning, blame). For each of these early codes, we explored sub codes that spoke to the mechanisms that made these actions possible — thinking about how students and facilitators were actively constructing disagreement, accountability, facilitation, etc.

Iterative review across data sessions allowed us to refine our research questions to focus on the nature of engagement in this context — drawing attention to nonverbal discourse in particular. This led to the reformattting of transcripts to privilege nonverbal action. Orienting to embodied interaction, we added layers of codes related to nonverbal features that appeared to be shaping the interaction. These codes fit into three categories (movement of artifacts, gesture, and gaze). Three episodes were selected that best represented each of these categories in action. Choosing such moments in which nonverbal engagement is well represented is particularly important for investigating how nuanced collaboration contributes to knowledge convergence and to deep learning. The findings will present examples of each of these discursive features and the nuanced functions they performed. These episodes speak to the nature of this small group’s collaborative work and the nonverbal actions used to move towards the construction of a JPSS.

Findings

We present three episodes that highlight discursive themes that emerged in our qualitative analysis: discrepancies, accountability and ownership, and parallel interaction. Each of the episodes highlights a theme that speaks to the construction of a JPSS and illuminates the nonverbal aspects of the small group’s collaborative work. As we trace these discourses, we map them to the group’s activity — working to understand how and when this group achieved a shared problem space. These findings provide a developing framework for discursive features that shape the construction of JPSS in our HCR curriculum.

Episode 1

As students called attention to and negotiated discrepancies throughout the HCR unit, a potential space was created for knowledge convergence. In the movement of artifacts across the table and verbal uptake and rejection of group members’ ideas, students performed accountability and ownership. This performance shed light on how the group was negotiating discrepancies, but it also illuminated norms around what it meant to “do STEM” in this space. As students coordinated their activity, the words they used and the movement of materials in the workspace shaped what they achieved. We conjecture that the interaction analyzed in the following two episodes can be labeled a JPSS, though it may not be one that was productive for knowledge convergence. This episode demonstrates that not all convergence is necessarily productive.

During the session in which the first episode takes place, students used maps they created of their classroom to consider a route that a robot could take in order to allow remote students to see the most important elements of the room. Each group in the class experimented with a “human robot” to develop specific directions they could translate into coded language to be used on a programming platform developed for this unit (http://robotmoose.com). As groups tested their human robot directions (e.g., “walk forward 10 steps”), they found ways to refine their directions to be increasingly specific for a wide audience (e.g., using units like “floor tiles,” “feet,” and “degrees”). Though the group put forth several ideas for units (tiles on the floor, the metric unit of a “foot”) they did not all use the same units in their individual written work. Thus, they were engaged in a parallel interaction where each individual was pursuing his or her own solutions (Barron, 2000). In this first episode, the discrepancies in the four students’ individual work come to the fore. In the activity that follows (Figure 3), the group’s productive negotiation of these spaces was impeded. This highlights unsuccessful JPSS as a precursor to this group’s successful construction of a JPSS. The episode began with student Karen’s attempt to check the work that she has been doing independently. Her groupmates (Chad, Rayna, and Evan) filled in their own individual robot direction worksheets (see Appendix A for transcript conventions). All student names have been changed.
In this episode, ownership and authority were performed in the positioning of artifacts. As Chad distributed worksheets (line 2), he foregrounded the separation of solutions — positioning each set of directions as belonging to one individual (Rayna). When Rayna rejected the paper that Chad handed her (line 3), she rejected personal ownership, as “everybody wrote on it.” This statement highlighted Rayna’s attention to the group as a community. At this point, Evan was positioned as an authority (line 1 and 6), likely because he was the student to physically walk thepaces while measuring the number of tiles for each move. His peers oriented to him and his work. The movement of artifacts across the table provided a visible representation of their negotiation of authority as they worked to develop a solution. In lines 6-8, the artifact movement across the table served as an emerging boundary object — passed between two participants, but not yet achieving its potential as a conduit for joint problem solving. All students continued to write on their own individual worksheets, deferring to the perceived group authority (Evan) as they copied his responses. We view this as a failed JPSS. In the episodes that follow, we explore more successful collaborative problem-solving experiences and what made them possible.

Episode 2

Episode 2 demonstrates how the group’s coordination of ownership and display of engagement continued across multiple levels. This extract occurred early in a class period when students were preparing to present their group’s robot design to the whole class. In the days leading up to this, students proposed individual robot designs and agreed upon a shared design for the group. This design was then submitted to be 3D printed as a small prototype by the research team. In these lines of interaction, we see how students negotiated authority in the group and discrepancies about one feature of the 3D-printed design. Throughout this interaction, students engaged in the developing and testing of solutions and improving phases of the design cycle. In this episode, nonverbal transcription sheds light on the role that nonverbal discourses played within the interaction (see Figure 4).

In line 1, Chad used gesture to reference the group’s design drawing on the table as he asserted that the group’s 3D printed model did not reflect what they represented in their sketch. Here, he called attention to the discrepancy between drawn fingers on their hand-like robot and the 3D printed rendering. This subtle gesture demonstrated Chad’s engagement with the group’s process and shared design history, indicating that he was working within a JPSS organized around that history. The movement functioned to orient the group to what they had already done (the design drawing) and where to go next (articulating their idea to the whole class). Rayna took up this orientation and contributed to the group’s coordination in line 2 — articulating what the design was “supposed to be like.” This statement of “supposed to be” asserted knowledge about the design. However, Rayna’s unfinished sentence and hand gesture complicated this action and created a space for refinement of the collective understanding (a possible JPSS). As she constructed her argument for facilitators positioned at the table, Rayna extended her own hand, using her body as a representation of what the robot hand should look like. She then abandoned this gesture, moving her hand to cover her eye (line 2). This movement functioned as a cue for others to jump in and participate. Rayna’s embodied gesture and her verbal reasoning both halted in line 2. The remainder of her sentence waited to be filled.

In line 3, Karen responded to this bid for participation from others by leaning back from the table and fixing her gaze on Rayna. This movement away from the design drawing shifted the responsibility back to Rayna. Rather than referencing the group’s designs or responding verbally, Karen leaned back and thus established herself as a
listener. This positioning further established Rayna’s ideas as representative of the collective group. In this moment, the potential JPSS seemed to narrow and disintegrate — centering again on one individual as responsible for problem solving. Here, we see how these students are moving apart and coming together as they use discourse and material objects to negotiate problem solving.

<table>
<thead>
<tr>
<th>Line</th>
<th>Name</th>
<th>Verbal transcript</th>
<th>Nonverbal transcript</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>See the design had fingers but the new one didn’t.</td>
<td>Gestures towards design drawing with right hand. This hand also holds a piece of the 3D printed model.</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
<tr>
<td>2</td>
<td>R</td>
<td>I mean like () it was supposed to be like--</td>
<td>Rayna extends her right hand in front of her chest and then moves it over her eyes.</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
<tr>
<td>3</td>
<td>K</td>
<td></td>
<td>Leans back from table to sit upright, gaze directed at Rayna.</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
<tr>
<td>4</td>
<td>Facilitator 2 (F2)</td>
<td>Like a hand like this?</td>
<td>Facilitator 2 arrives at the table.</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
<tr>
<td>5</td>
<td>Facilitator 1 (F1)</td>
<td>Like a hand like this?</td>
<td>Facilitator 1 holds up her hand up with palm facing the table, mirroring the beginning of Rayna’s gesture.</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>:=I know but you told me to do it like that</td>
<td>Moves left forearm across table with palm facing himself towards the design drawing that sits in front of Rayna and Karen, then back to rest on the table in front of him</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
<tr>
<td>7</td>
<td>R</td>
<td>Okay, I told him to do it like that=</td>
<td>Raises right forearm from table, moves right hand in a flipping motion with palm moving away from her body</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
<tr>
<td>8</td>
<td>F2</td>
<td>=You don’t like the square ones like that?</td>
<td>Facilitator 2 points to drawing</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
<tr>
<td>9</td>
<td>E</td>
<td></td>
<td>Leans in towards drawing in front of Chad as facilitator begins speaking</td>
<td><img src="image" alt="Image of Rayna extending her right hand in front of her chest and then moving it over her eyes." /></td>
</tr>
</tbody>
</table>

In line 5, a facilitator mirrored the beginning of Rayna’s gesture. This movement functioned as a piece of the facilitator’s role — ending Rayna’s sentence via embodied movement and guiding the group’s work with a question: “Like this?”

Lines 6-7 include examples of gestures that speak to accountability and ownership in this small group. In line 6, Chad gestured toward the group’s design drawing as he verbally placed blame for the state of the group’s design (“you told me to”). This gesture drew attention to the shared artifact and evidence at the table — contributing to Chad’s construction of accountability as the group negotiated a discrepancy. Rayna used her own gesture to dismiss in the following line — flipping her wrist outward as she declared: “Okay, I told him to do it like that.”
Facilitator 2 then referenced the drawing as he mediated the group’s interaction. This embodied discourse from an authority figure validated Chad’s move to make the artifact relevant as evidence in the argument.

Finally, in line 9, we see the action of leaning made relevant for the second time in this exchange. Evan, who had not yet spoken, leaned forward in his chair—craning his neck towards the design drawing. This subtle gesture demonstrated his peripheral engagement in a JPSS. Throughout this extract, we see how gesture was used to make shared artifacts relevant in group reasoning and how body positioning (e.g., leaning) can be linked to attention and peripheral participation (line 9) and division of labor (line 3). We also see how gesture was used as a facilitation strategy—reorienting the group to their shared design. This re-orientation might be leveraged as a way to help groups recognize and reflect upon their JPSS in real time.

Throughout this episode, a JPSS was created through nuanced performances of accountability and ownership. These discursive performances served to construct a shared space for problem solving. While we do not see deep learning and knowledge convergence at this stage, this demonstration of a JPSS acts as a precursor. In the episode that follows, we move beyond the construction of accountability and ownership to see how discrepancy was negotiated by this small group and the class as a whole.

### Episode 3

Throughout the process of creating and coding nonverbal transcripts for the analysis featured here, pointing emerged as a significant feature of group interaction. In preliminary coding, we identified three functions of pointing: accountability (e.g., directing another student to respond), attribution of blame (e.g., pointing that highlights a discrepancy), and evidence (e.g., pointing to artifacts that support a claim). Episode 3, presented below, gives an example of pointing that provided evidence for deep engagement with design reasoning (Figure 5). Prior to the beginning of this episode, a classroom facilitator (Facilitator 2) discussed the 3D printing process for a prototype of the robot that the group had drawn. He went on to describe some of the decisions he had to make when he translated the group’s drawing into 3D form. The group noticed aspects they intended did not come through clearly in the 3D model. They used pointing as a means of coordinating discussion around these features, attempting to create a shared understanding of the problem.

<table>
<thead>
<tr>
<th>Line</th>
<th>Name</th>
<th>Verbal transcript</th>
<th>Nonverbal transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F1</td>
<td>So what’s this one?</td>
<td><strong>Pointing to a feature of the design</strong></td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td></td>
<td>Gaze directed at design drawing and 3D model</td>
</tr>
<tr>
<td>3</td>
<td>F2</td>
<td>take it home can.]</td>
<td>Facilitator is finishing the sentence “anyone who wants to take it home can”—overlapping from interaction immediately preceding this episode.</td>
</tr>
<tr>
<td>4</td>
<td>R</td>
<td>Oh that’s a cup holder</td>
<td><strong>Picks up the 3D model and points to a feature</strong></td>
</tr>
<tr>
<td>5</td>
<td>K, C, E</td>
<td></td>
<td>Gaze directed at Rayna</td>
</tr>
<tr>
<td>6</td>
<td>R</td>
<td>It’s supposed to have like a bottom on it.</td>
<td></td>
</tr>
</tbody>
</table>

---

*Figure 5. Episode 3*
In lines 1 and 4 of this interaction, pointing directed attention to features of the group’s 3D modeled design. Facilitator 1’s pointing called attention to a specific feature of the group’s 3D printed design model (line 1). Evan directed his gaze to this feature — orienting to the move made by the facilitator to achieve joint attention (line 2). Rayna physically handled the 3D model and pointed to the highlighted feature (line 4). She explained that it was intended as a cup holder. Here, Rayna gave voice to the collective ideas of the group. With this move, she used the physical model at the table and the gesture of pointing to support her explanation. As she gestured and explained, the other three members of the group oriented to her, as demonstrated by their gaze (line 5).

<table>
<thead>
<tr>
<th>Line</th>
<th>Name</th>
<th>Verbal transcript</th>
<th>Nonverbal transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>C</td>
<td>(. ) [Rayna], you said you can’t pick up a cup without fingers.</td>
<td>Tags table with left hand on design drawing. E.R.K gaze directed at C’s hand. C extends left fingers out from palm, moves hand up from table and then down as he says “fingers.”</td>
</tr>
<tr>
<td>8</td>
<td>K</td>
<td>It did have fingers on it. And it was the side view.</td>
<td>Holds her hand vertical in front of her with left elbow on the table.</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>It had this on front, side, and top.</td>
<td>C mirrors K’s hand motion. K lowers her head between her elbows on the table, directs gaze at R.</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>I mean it had (. ) that’s all it was.</td>
<td>C moves the design drawing on the girls’ side of the table closer to his spot at the table.</td>
</tr>
<tr>
<td>11</td>
<td>K</td>
<td>Okay, it doesn’t matter.</td>
<td></td>
</tr>
</tbody>
</table>

Following this interaction, the physical model of the group’s robot was passed between group members as they further explained their intended design. Here, students participated in a shared activity structure that made use of tools (the robot model and design drawings at the table) to communicate their collective robot design goals. This tool use helped the group to address issues that were not immediately visible in their prior drawing, and it also appeared to support students’ developing understanding of the need to iterate using physical prototypes within a design cycle. In this case, the physical robot prototype served as a boundary object for the group. In the following interaction, a later portion of episode 3, the students addressed discrepancies in their design ideas using pointing and physical artifacts (see Figure 6).

Throughout this connected episode, Chad worked to re-orient his group members to their design drawing and to come to consensus about their robot’s appearance. Again, the group refined their design ideas. As he pointed to the design drawing in line 7 and enacted robot movement, Chad argued for the inclusion of fingers in their final robot design. He also referenced Rayna’s idea — prompting the group to remember their shared decisions. This orientation to a shared design space and history served to construct a JPSS. In line 8, Karen used her hand to contribute. Here, she used physical gesture to show Chad what she and Rayna were depicting in their design drawing — a flat hand with touching fingers. Chad mirrored this gesture in line 9. As Chad and Karen mimicked
one another, they appeared to be constructing a shared and embodied means of representing their robot design. Chad further supported the construction of this shared understanding by bringing the original design drawing back into the conversation (line 10). In this interaction, the students experienced real issues in designing robots that will complete specific tasks (e.g., grasping). This authentic experience helped them to recognize the limits of static images in the iterative brainstorming phase of the design cycle.

For the purposes of this analysis, we claim that pointing is a meaningful feature of engagement and a site for co-construction of understanding. We provide one example of how pointing functioned to direct attention and provide evidence in this small group’s work with human-centered robotics. We conjecture that this particular gesture is important for groups’ communication of design and construction ideas, to command joint attention, negotiate discrepancies, and to make connections — thus serving to build toward a JPSS.

Discussion

In examining these episodes in a middle school HCR context, our analysis oriented to the functions of discrepancy, ownership and accountability, and parallel interaction in one group’s discourse. First, we showed how the performance of ownership and accountability can both support and impede joint problem solving. Across these episodes, we showed how the HCR context provided opportunities to create JPSS (even when not realized). The curriculum provided physical materials and a designed problem that allowed students to negotiate iteratively designed artifacts, the ownership of which could change over time. These curricular aspects, which support the creation of JPSS, can be leveraged further to deepen collaborative engagement with the engineering design cycle and the socio-technical systems that are critical aspects of HCR. Building upon the productive use of physical robot prototypes and shared written artifacts as boundary objects, we might integrate more opportunities for students to physically move ideas across the table. Tools and activities such as collage boards for design ideas and collaborative word documents can promote embodied argumentation and scaffold the creation of a shared space for meaning making. Students might be encouraged to use movement and gesture to explain — modeling ways that embodied and verbal actions can contribute to the construction of shared understanding. Furthermore, video reflection might be used with instructors to scaffold active noticing of embodied engagement. Attending to joint interaction and gesture can serve as an indicator of group engagement.

Throughout our analysis, we were particularly interested in how group collaboration and problem-solving occurred. As students engaged with their HCR projects, they moved apart and came together. Students were not always in a JPSS, but over time, the ability to imagine the others’ perspective became richer so that even in parallel work it was possible to attend to the perspective of the group. In the work presented here, we are interested in the social tensions and triumphs of small group work, considering how group dynamics and engagement are constructed and constructive in problem solving as well as how we can design for them. As students moved towards the construction of a JPSS, they came together more than they moved apart. We were able to see this coming together as we tracked verbal and nonverbal collaborative contributions. This analysis sets the stage for future work exploring JPSS, collaboration, and deep learning.

Conclusion

This research made a case for the importance of nonverbal engagement in collaborative interaction and presented a preliminary framework of four specific features that were significant in one small group’s collaborative interaction in a human-centered robotics curriculum. This framework includes embodied actions of leaning, the movement of artifacts, pointing, and gaze. These embodied elements functioned to position authority and accountability, direct attention, and provide support for verbal reasoning. These embodied actions are highlighted as aspects of deep engagement, and they contribute to existing literature on engagement and what it looks like in collaborative learning environments (e.g., Barron, 2000; Sinha et al., 2015). Tracing embodied actions throughout a series of episodes, we’ve demonstrated how they help us to “see” JPSS taking shape. Though these spaces are difficult to pin down and represent, we have highlighted how nonverbal actions taken by group members serve to create a shared mental and physical environment for meaning making.

This research has also shown that not all convergence was productive for group work. As students leaned and directed gaze, they communicated their participation and expectations about participatory patterns in group work. As they moved artifacts and pointed, they used material objects as artifacts to think with (Papert, 1980). These actions played out over time to weave the group’s collaborative fabric as they wrestled with the challenge of engineering design practices. Embodied action supported students’ brainstorming and iterative design. As they
displayed and negotiated ownership and accountability in the problem-solving process highlighted here, students engaged with authentic STEM practices — perhaps beginning to see themselves as STEM learners in a new light. An open question remains as to whether engaging in these more robust practices leads learners to appreciate them over time, suggesting that we need to continue to consider the relationship between deep engagement and deep learning.

This research demonstrated how collaboration in verbal and non-verbal dimensions was an important part of students’ design process, and thus that embodiment was central to students’ understanding of robotics design. As students engaged in embodied ways, they made connections and developed more robust collaborative learning practices. In this sense, they were engaging deeply. From our sociocultural perspective, this kind of deep engagement and participation is deep learning. Opportunities like these for deep STEM engagement are important for the development of STEM interest and collaborative practices. For non-dominant populations in STEM who often find it difficult to identify themselves in STEM activities and careers, experiences that prompt exploration of diverse embodied and verbal discourses may help to re-define what STEM can and should look like.

Acknowledgements

This research was funded by the National Science Foundation ITEST grant # 1433414. We thank Dr. Joshua Danish for comments on an earlier draft and all undergraduate and graduate student researchers who contributed to data collection and analysis sessions. These students include AnnaRose Girvin, Joey Huang, Stella Huang, Charles Mahoney, Miranda Meade, Haley Molchan, and Benjamin Oistad.

References


Appendix A. Transcript conventions (adapted from Jefferson, 2004)

Transcript conventions

[ A single left bracket notes the beginning of overlapping talk
] A single right bracket notes the end of overlapping talk
= An equal sign notes latching talk (no gap between turns)
(1) Numbers in parentheses note pauses. (1) notes a pause of ~ 1 second
(. ) A period in parentheses notes a small pause between utterances
__ Italic __ Underlined speech notes an emphasis on the underlined word(s)

Student pseudonyms and table positions

Rayna  Chad
Karen  Evan

Front of classroom/bottom of transcript images
Deep Learning towards Expertise Development in a Visualization-based Learning Environment

Bei Yuan1,2, Minhong Wang1*, Andre W. Kushniruk3 and Jun Peng4

1KM&EL Lab, Faculty of Education, The University of Hong Kong, Hong Kong // 2Party School of Zhongshan Municipal Party Committee, China // 3School of Health Information Science, University of Victoria, Canada // yuanei1@hku.hk // magwang@hku.hk // andrek@uvic.ca // u3002116@connect.hku.hk

*Corresponding author

ABSTRACT

With limited problem-solving capability and practical experience, novices have difficulties developing expert-like performance. It is important to make the complex problem-solving process visible to learners and provide them with necessary help throughout the process. This study explores the design and effects of a model-based learning approach implemented in a web-based learning environment that not only allows learners to capture and reflect on their problem-solving process in visual formats but also helps them to identify the gap between their performance and that of the expert for effective reflection and improvement. The proposed approach attempts to utilize expert knowledge to transform open-ended problem-solving experience into a systematic and deliberate effort towards expertise development. Twenty-five medical students participated in the study by using the proposed learning environment to complete a number of diagnostic problem-solving tasks. The results show that the approach positively affects students’ achievements in subject knowledge, problem-solving performance, and perceptions and motivation for learning in the proposed learning environment.

Keywords

Problem solving, Expertise development, Professional development, Visualization, Computer-based learning environment, Model-based learning and instruction

Introduction

As a form of constructivist learning, problem solving has received wide attention in educational practice, especially in complex and ill-structured domains such as scientific inquiry and medical education (Jonassen, 1997). Problem-solving experience can help learners to develop critical thinking, communication, and problem-solving skills as well as improve the construction of knowledge (Hmelo-Silver, 2004). Given that learning with real-world problems is constrained in classroom settings, computer-based environments have been increasingly explored as a way to support learning through problem solving in virtual environments. Computer-based learning environments have clear advantages in affording flexible access to learning resources, on-demand delivery of learning programs, flexible communication with others, and more importantly computer-based learning support.

However, effective learning through problem solving is difficult to realize in both classroom and computer-based settings. Solving a real-world problem often involves a sophisticated process of understanding the problem, linking abstract knowledge to problem information, and applying relevant methods and strategies to solve the problem. Learning in such contexts can generate a heavy cognitive load for learners (Kirschner, Sweller, & Clark, 2006) that instructors or experts often underestimate, as for them many of the requisite processes have become largely automatic or subconscious with experience. With limited abilities to capture the complex problem-solving process, many learners fail to adequately engage in authentic task experience and achieve desired learning outcomes.

Learning through problem-solving experience has been widely adopted in medical education by way of problem-based learning curricula, case-based sessions, and internship programs. Several reviews of the literature have shown that problem-based learning improves students’ reasoning and communication skills, fosters their abilities to cope with uncertainty, and empowering self-directed learning (Albanese & Mitchell, 1993; Dochy, Segers, Van den Bossche, & Gijbels, 2003; Hartling, Spooner, Tjosvold, & Oswald, 2010; Koh, Khoo, Wong, & Koh, 2008; Neville, 2009). At the same time, researchers have reported inconclusive and inconsistent findings on the superiority of problem-based learning over conventional instructions, mainly in systematic construction of subject knowledge and the development of efficient reasoning process (Coderre, Mandin, Harasym, & Fick, 2003; Patel, Yoskowitz, Arocha, & Shortliffe, 2009). A major concern is that completing a problem-solving task such as clinical diagnosis involves complex cognitive processes in the search for problem information about multiple aspects, integrating the problem information with subject-matter knowledge, and reasoning with interactive components to analyze the problem (Delany & Golding, 2014; Wang, Wu, Kinshuk, Chen, & Spector, 2013).
Previous studies report that scaffolding supports learning in complex task situations mainly by using prompts or tips to bring learners’ attention to important issues or by decomposing a complex task into a set of main actions or key questions (Ge, Chen, & Davis, 2005; Hmelo-Silver, Duncan, & Chinn, 2007). Recent research has highlighted that when working with complex problems, making thinking visible with the support of visualization-based learning technology is important (Wang & Jacobson, 2011). Externalization of complex cognitive processes or mental models promotes deep learning and improves learning outcomes in inquiry learning and problem-solving contexts (Gijlers & de Jong, 2013; Linn, 2000; Suthers, Vatrapu, Medina, Joseph, & Dwyer, 2008; Van Bruggen, Kirschner, & Jochems, 2002; Wu, Wang, Grotzer, Liu, & Johnson, 2016). In doing so, visual representations and graphic forms play an important role in representing complex thinking and cognition in flexible ways.

In addition to enabling learners to capture their cognitive process or mental model for deep thinking and self-reflection, there is a need to explore how learners’ performance can be further improved by allowing them to reflect on the gap between their process and that of the expert so as to promote expertise development. Research on expertise development has revealed that desired learning outcomes in problem-solving contexts cannot be achieved through a mere accumulation of experience, but require systematic and deliberate effort with expert support (Ericsson, 2008; Jarodzka, Scheiter, Gerjets, & van Gog, 2010). Further to our prior studies on scaffolding thinking and reflection by externalizing and visualizing a set of cognitive elements in problem-solving contexts (Wang et al., 2013; Wu & Wang, 2012; Wu et al., 2016), the present study aims to explore how novice performance can be further improved through the utilization of expert knowledge and visualization-based learning facilities.

This study uses a visualization-based learning environment design that helps medical students to capture the complex process of diagnostic problem solving with the support of expert feedback. The design features model-based learning, which attempts to enable students to capture and reflect on their problem-solving process and improve their performance by identifying the gap between their process and that of the expert. The research questions (RQs) of the study are specified as follows:

**RQ1:** How can a computer-based learning environment be designed to help medical students capture the complex process of solving a diagnostic problem with expert support?

**RQ2:** What are the effects of the proposed approach on student learning in a diagnostic problem-solving context?

### Theoretical framework

Learning through problem solving positions learners in real-world problem contexts, helping them to develop critical thinking and problem-solving skills as well as consolidate and extend content knowledge. Learning in problem-solving contexts is supported by situated cognition theory (Brown, Collins, & Duguid, 1989) and situated learning theory (Lave & Wenger, 1991). The two theories share a common view that situation and cognition are interdependent. Cognition is a process that occurs in physical and social contexts where knowledge is created and applied. Problem-solving experience is also recognized as an integral part of expertise development (Ericsson, 2008; Schmidt, Norman, & Boshuizen, 1990).

Given that open-ended exploration with complex problem-solving tasks can overburden learners, the provision of scaffolding or support to learners has been widely recognized as an important part of learning in such situations (Hmelo-Silver et al., 2007). The use of scaffolding to support student learning with complex problems is aligned with the cognitive apprenticeship model (Collins, Brown, & Holum, 1991), which claims that carrying out a complex task usually involves implicit processes. It is critical to make such processes visible for novices to observe and practice, and to provide them with expert help. At the same time, when scaffolding learning in complex situations, it is important that the scaffolding or support does not undermine the open-endedness of the task and individual endeavor.

Making complex tasks and thinking processes visible can be linked with model-based learning and instruction, i.e., the use of mental models to uncover the cognitive processes and architectures to gain insight into the nature of complex problem solving (Greca & Moreira, 2000; Seel, 2003). A mental model is “what people really have in their heads and what guides their use of things” (Norman, 1983). Effective learning in problem-solving contexts requires the externalization of the implicit mental models that are associated with sequences of actions and the underlying knowledge in complex problem-solving processes (Bradley, Paul, & Seeman, 2006). Model-based learning and instruction has two different modes: self-guided and expert (Seel, 2003). In the self-guided mode, students are expected to develop their own mental models with little support or guidance, which is more suitable
for well-structured problem solving, or when students have profound knowledge and experience in a given domain. In the expert mode, experts’ mental models are externalized as support and a guide to help students solve complex problems or accomplish learning tasks. It is more suitable for ill-structured problem solving, or when students have limited prior knowledge and experience.

Methods

This study uses the design-based method because it aims to develop a computer-based learning environment for diagnostic problem solving and examine its effects on student learning. As a systematic and scientific methodology, design-based research involves iterative analysis, design, development, and implementation to create and evaluate innovative interventions to solve discovered problems (The Design-Based Research Collective, 2003; Reeves, Herrington, & Oliver, 2004; Wang & Hannafin, 2005). It has been employed to the design of educational interventions to promote sustained and practical development in technology-enhanced learning environments (Dede, 2004). It is particularly suitable when complex and ambitious educational reform policies are ill specified and the implementation process is uncertain (Wang, Vogel, & Ran, 2011).

Informed by relevant learning theories and instructional models, a computer-based, expert-supported learning environment has been designed to help medical students to capture the complex process of diagnostic problem solving to improve their problem-solving expertise (Yuan, Wang, Kushniruk, & Peng, 2016). An empirical study has been conducted with medical students to determine the effects of the proposed approach by examining students’ learning outcomes using the proposed approach.

Glaucoma diagnosis was chosen as the learning subject because it is a part of the learning content of general courses in medical schools and is considered to involve ill-structured problem solving. Two domain experts with over ten years of clinical and academic experience in glaucoma diagnosis and treatment participated in this study. They supported the preparation of clinical cases and the assessment of learning outcomes. A medical teacher with years of teaching experience in a public medical school participated in the study to support the arrangement of learning activities with students.

Seven clinical cases were used for the study. Five of these were used for learning tasks, and the other two for assessing student performance before and after the study. All of the cases were selected and adapted from authentic clinical cases by the experts. The selection criteria were that the cases should be representative and clearly presented and have referenced solutions validated by the experts. The reference solutions were used to assess student performance and provide feedback to students during the task process.

Proposed learning environment

Glaucoma Diagnosis & e-Learning System (GDeL), a computer-based learning environment used to support students’ learning and expertise in glaucoma diagnosis, has been developed and includes the following main functions.

**Exploratory problem-solving context**

The exploratory problem-solving context allows learners to work with a number of simulated diagnostic problems. After selecting a case in the system, the learner can view a primary description of the patient’s background information (e.g., age, gender, and medical history) and chief complaint. Based on the initial information, the student forms an initial assessment of the case. Moreover, the learner can conduct clinical examinations of the patient to obtain further information, as shown in Figure 1. For each selected clinical examination, the learner can view the examination results (in the form of laboratory data, images, and brief inspection reports) and make intermediate judgments based on the results. After obtaining adequate information via several rounds of examination and judgment, the student makes a diagnostic conclusion for the case.
Visualization of the problem-solving process

For each case, the diagnostic problem-solving process performed by the learner can be captured by the system and shown in a flowchart (see Figure 2). A diagnostic flowchart covers the initial information about the patient, clinical examinations selected for the patient, intermediate judgments made upon receiving the examination results, and a diagnostic conclusion. By externalizing the diagnostic process in a visual format, the system allows the learner to review and reflect on his/her task process and performance.

Expert support

The learner can practice with the same case repeatedly. After each diagnosis, the learner can view the feedback about his/her performance that the system generates by comparing his/her performance with that of the expert, e.g., “Your process missed some examinations.” The feedback focuses on three key aspects of the performance, i.e., selection of clinical examinations, intermediate judgment based on the examinations, and diagnostic conclusion. The feedback allows the learner to recognize whether he/she has missed necessary examinations or judgments or included unnecessary or inappropriate examinations, judgments, or conclusions. The feedback is not concerned with the sequence of actions (e.g., the order of the clinical examinations), which involves certain flexibility. The performance in collecting initial information about the case was not taken into account, as all of the learners were provided with the same original information for each case, which was very limited and useful for diagnostic analysis.

Once the overall similarity in task performance between the student and expert reaches 60% or more, or the learner has tried to solve the case 10 times, the learner can view the expert’s diagnostic process, together with the expert’s summary of the key points and common errors in case diagnosis (see Figure 3).
Construction of knowledge in mental map

The system also includes a diagram tool, which allows students to build a mental map to externalize the subject-matter knowledge underlying the diagnostic processes. Students are encouraged to reflect on the knowledge underlying all cases at the end of the study and represent it in a mental map in a flexible way (see Figure 4).

The overall design of the GDeL system is aligned with the six strategies proposed in the cognitive apprenticeship model: modeling, coaching, scaffolding, articulation, reflection, and exploration. The learning environment features an exploratory problem context for exploration with authentic problems. The complex process of a diagnostic task is scaffolded by highlighting key actions in clinical examination, intermediate judgment, and diagnostic conclusion. Learners are able to capture and reflect on their problem-solving process. Learners receive coaching via feedback on their performance, which is associated with the expert model that helps learners to determine the gap between their performance and the expert performance. Based on experience over multiple cases, learners are encouraged to articulate the knowledge underlying the problem-solving process.
Participants

Twenty-five Year 4 students from a public medical college participated in the study. The participants had basic knowledge and skills needed for clinical reasoning and diagnosis.

Learning task

The participants used the proposed learning environment to complete five simulated glaucoma cases. The task process was designed to mirror clinical encounters in that learners were given incomplete information about a problem and had to collect further information by selecting clinical examinations and making intermediate judgments based on the examination results in several rounds before reaching a diagnostic conclusion.

Students could diagnose a case more than once. After each diagnosis, they could review their diagnostic process in a visual format, and receive feedback about the degree of similarity between their performance and that of the expert. Moreover, as noted earlier, in due course the learner was able to view the expert’s diagnostic process and summary of the key points and common errors in case diagnosis. After accomplishing all of the cases, the learner was encouraged to externalize the subject-matter knowledge underlying the diagnostic processes of all of the cases in a mental map using a diagram tool provided by the system.

Procedure

The learning program lasted for six weeks. In the first week, students signed the consent forms for their participation in the study. A questionnaire survey was then administered to the participants to collect their demographic data. A pre-test was also administered to assess their subject knowledge, and they were provided with a clinical case to assess their diagnostic problem-solving performance. Students were then given a face-to-face, one-hour introduction on how to use the system to perform diagnostic problem solving and reflective learning with clinical cases. A sample case was used to demonstrate the learning process, and to enable students to practice and become familiar with the learning environment.

Students started their self-directed learning in the second week. They were asked to complete five simulated cases within four weeks. They were advised to pace themselves, and spend three to four hours per case. During the task period, there was no teacher involvement except for assistance with technical problems. Students could use online forums for flexible discussion and communication with other participants.

In the sixth week, a post-test was arranged with students to assess their knowledge achievement, and a clinical case was administered to assess their diagnostic problem-solving performance. A questionnaire survey was administered to collect students’ perceptions of the learning environment and the cognitive strategies supported by the learning environment, in addition to students’ motivation for learning. A semi-structured written interview was used to collect students’ comments and feedback about the learning program by requiring students to write the responses on the paper.

Measures

The learning outcomes were assessed using subject knowledge tests and diagnostic problem-solving tasks before and after the learning program, and a questionnaire survey and semi-structured written interview at the end of the study.

Pre-test questionnaire

The pre-test questionnaire was used to collect students’ gender and self-assessment of their computer skills (very poor, poor, intermediate, good, very good) and intention to use computer-assisted learning applications (from strongly disagree to strongly agree).
Post-test questionnaire

The post-test questionnaire was used to collect students’ perceptions of the learning system and the cognitive strategies supported by the learning system, and students’ motivation for learning using the proposed approach. The survey used a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The items measuring learner perceptions of the learning environment were adapted from widely used information technology acceptance instruments (Arbaugh, 2000; Davis, 1989). These items measure students’ perceptions of the learning system in terms of usefulness, ease of use, and intention to use. Examples of the items include “The system is useful for my learning,” “The system is easy for me to use,” and “I have an intention to use the system.”

The items measuring learners’ perception of the cognitive strategies supported by the learning environment were adapted from the instrument evaluating clinical instruction and learning environment in the clinical practice setting (Stalmeijer, Dolmans, Wolfhagen, Muijtjens, & Scherpbier, 2010). The instrument consisted of six subscales: modeling, coaching, scaffolding, articulation, reflection, and exploration. The validity and reliability of the instrument has been well established (Stalmeijer, Dolmans, Wolfhagen, Muijtjens, & Scherpbier, 2008; 2010). Examples of the items include “The system facilitates my reflection of the problem-solving process,” “The system provides me the opportunity to exhibit my understanding of domain knowledge,” and “The expert advice provided by the system is helpful for my study.”

The instrument for evaluating motivation developed by Keller (2010) was adapted to measure student motivation for learning using the proposed approach. The instrument involves four scales including attention, relevance, confidence, and satisfaction. Examples of the items include “The learning program inspires my curiosity,” “The learning program is related to my expectations and goals,” “I feel confident that I will do well in this learning program,” and “I feel satisfied with what I learnt from this program.”

Knowledge tests

The knowledge achievement made by students was assessed before and after the study using two traditional knowledge tests (pre- and post-test). Different questions of similar difficulty were used for the two tests. Each test comprised 10 single-choice, 10 multiple-choice, and 10 true-or-false questions. The scores ranged from 0 (incorrect) to 1 (full credit) for each question, with a test range of 0 to 30 rescaled to the range between 0 and 1. All of the questions were selected from the question bank of a medical school. The validity and appropriateness of the test questions and reference answer were endorsed by the two experts. Each test was completed within 45 minutes.

Problem-solving tasks

The problem-solving performance of the participants was assessed before and after the study using two cases at the same level of difficulty. The assessment was based on the degree of similarity between the student’s and expert’s performance as reflected in their diagnostic records. The assessment focused on three aspects: (1) selection of clinical examinations to collect further information, (2) intermediate judgments based on examination results, and (3) diagnostic conclusion. The performance in collecting the initial information about the case was not taken into account, as all of the learners were given the same original information for each case, which was very limited and useful for diagnostic analysis.

The performance in each of the three aspects was assessed based on the number of valid items, unnecessary items, and missing items in the diagnostic record, defined and specified as follows.

• Valid items are the elements present in the expert’s diagnostic record as well as in the learner’s diagnostic record.
• Unnecessary items are the elements present in the learner’s diagnostic record, but not in the expert’s diagnostic record.
• Missing items are the elements present in the expert’s diagnostic record, but missing in the learner’s diagnostic record.

Tversky’s (1977) contrast formula was used to measure the degree of similarity between the student’s performance and the expert’s performance in each of the three aspects, such that performance in each aspect is determined by the following formula.
The average score of the three scales represents the overall problem-solving performance. Therefore, the similarity scores ranged from 0 (indicating the diagnostic records are totally different) to 1 (indicating the diagnostic records are identical).

Two assessors who were blind to student identification and test information (i.e., pre- or post-test) appraised the diagnostic records independently. One rater graded the test papers and diagnostic records with reference to expert solutions, while the other checked and confirmed the grades. Any disagreements arising during the process were resolved through discussion.

Semi-structured written interview

The interview was used to collect students’ comments and feedback on the learning program by requiring students to write the responses on the paper. The interview included two open-ended questions: (1) the strengths and weaknesses of the learning program and (2) suggestions for improving the learning program.

Results

All 25 students (12 males, 13 females) completed the study. Most had between intermediate (48%) and good (40%) computer skills, and none self-reported as poor computer users. Consistent with this, most of the participants showed a positive intention (agree 56%, strongly agree 32%) to use computer-assisted learning applications.

Questionnaire survey

Perceptions of the learning environment

The participants reported positive perceptions of the learning system in terms of its usefulness (Mean = 4.08, SD = 1.11), the ease of use of the system (Mean = 4.14, SD = .92), and their intentions to use it (Mean = 4.06, SD = 1.01). An internal consistency analysis using Cronbach’s alpha confirmed that all of the subscales were reliable (.93 for usefulness, .91 for ease of use, and .94 for intention to use).

Perceived cognitive strategies

The participants had positive perceptions about the cognitive strategies supported by the learning environment in terms of modeling (Mean = 4.21, SD = .94), coaching (Mean = 4.21, SD = .93), scaffolding (Mean = 4.16, SD = .99), articulation (Mean = 4.13, SD = .94), reflection (Mean = 4.09, SD = .90), and exploration (Mean = 4.08, SD = .90). An internal consistency analysis using Cronbach’s alpha confirmed that all of the subscales were reliable (.97 for modeling, .97 for coaching, .96 for scaffolding, .96 for articulation, .95 for reflection, and .96 for exploration).

Table 1. Correlations between perceptions of the learning system and perceptions of cognitive strategies (n = 25)

<table>
<thead>
<tr>
<th></th>
<th>Modeling</th>
<th>Coaching</th>
<th>Scaffolding</th>
<th>Articulation</th>
<th>Reflection</th>
<th>Exploration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>0.812**</td>
<td>0.895**</td>
<td>0.903**</td>
<td>0.853**</td>
<td>0.784**</td>
<td>0.855**</td>
</tr>
<tr>
<td>Ease of use</td>
<td>0.921**</td>
<td>0.951**</td>
<td>0.916**</td>
<td>0.861**</td>
<td>0.862**</td>
<td>0.889**</td>
</tr>
<tr>
<td>Intention to use</td>
<td>0.880**</td>
<td>0.932**</td>
<td>0.907**</td>
<td>0.854**</td>
<td>0.897**</td>
<td>0.882**</td>
</tr>
</tbody>
</table>

Note. **p < .01.

The correlations between the learners’ perceptions of the overall learning environment and their cognitive strategies were analyzed using Spearman correlation coefficients. As shown in Table 1, learners’ perceptions of the GDeL environment were significantly correlated with their perceptions of the cognitive strategies supported by the GDeL environment. In particular, the scaffolding strategy supported by the system was most highly correlated with the perceived usefulness of the GDeL system, while the coaching strategy supported by the
system was most highly correlated with the perceived ease of use of the system and intention to use it for learning.

**Motivation to learn**

The participants had a clear motivation to learn using the proposed learning environment, as shown by the scales of attention (Mean = 4.11, SD = .98), relevance (Mean = 4.12, SD = .96), confidence (Mean = 4.13, SD = .96), and satisfaction (Mean = 4.08, SD = .99). As shown by the data, students had a strong interest in learning using the proposed approach, and felt that the learning program was highly relevant to their expectation. Moreover, they felt confident during the learning program, and were satisfied with the learning experience. An internal consistency analysis using Cronbach’s alpha confirmed that all of the subscales were reliable (.96 for attention, .96 for relevance, .97 for confidence, and .97 for satisfaction).

**Knowledge tests**

The students had significantly improved subject knowledge after completing the learning program (Pre-test: Mean = .48, SD = .08; Post-test: Mean = .60, SD = .09; p < .001).

**Problem-solving tasks**

As shown in Table 2, students made significant progress in all subscales of problem-solving performance after the learning program. From pre-to post-test, the mean value increased from .38 to .67 on selection of clinical examination to collect further information, from .18 to .54 on intermediate judgment based on examination results, and from .08 to .40 on diagnostic conclusion. Consequently, the mean value of overall problem-solving performance increased from .21 to .54. Their performance on intermediate judgment based on examination results improved the most, as depicted in Figure 5.

<table>
<thead>
<tr>
<th>Scales</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of valid items</td>
<td>1.76</td>
<td>.72</td>
<td>3.36</td>
<td>.86</td>
<td>-7.16</td>
</tr>
<tr>
<td>Number of unnecessary items</td>
<td>4.32</td>
<td>1.35</td>
<td>2.80</td>
<td>1.66</td>
<td>3.48</td>
</tr>
<tr>
<td>Number of missing items</td>
<td>1.24</td>
<td>.72</td>
<td>.64</td>
<td>.86</td>
<td>2.68</td>
</tr>
<tr>
<td>Similarity to expert</td>
<td>.38</td>
<td>.15</td>
<td>.67</td>
<td>.19</td>
<td>-5.41</td>
</tr>
<tr>
<td>Intermediate judgment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of valid items</td>
<td>.80</td>
<td>.76</td>
<td>2.60</td>
<td>1.00</td>
<td>-8.05</td>
</tr>
<tr>
<td>Number of unnecessary items</td>
<td>5.28</td>
<td>1.54</td>
<td>3.72</td>
<td>1.82</td>
<td>3.21</td>
</tr>
<tr>
<td>Number of missing items</td>
<td>1.24</td>
<td>.72</td>
<td>.64</td>
<td>.86</td>
<td>2.68</td>
</tr>
<tr>
<td>Similarity to expert</td>
<td>.18</td>
<td>.17</td>
<td>.54</td>
<td>.20</td>
<td>-7.32</td>
</tr>
<tr>
<td>Diagnostic conclusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of valid items</td>
<td>.08</td>
<td>.28</td>
<td>.40</td>
<td>.50</td>
<td>-2.87</td>
</tr>
<tr>
<td>Number of unnecessary items</td>
<td>.92</td>
<td>.28</td>
<td>.60</td>
<td>.50</td>
<td>2.87</td>
</tr>
<tr>
<td>Number of missing items</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Similarity to expert</td>
<td>.08</td>
<td>.28</td>
<td>.40</td>
<td>.50</td>
<td>-2.87</td>
</tr>
<tr>
<td>Overall</td>
<td>.21</td>
<td>.14</td>
<td>.54</td>
<td>.18</td>
<td>-6.46</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05; **p** < .01; ***p*** < .001.
Semi-structured interview

The comments shared by students in their responses are presented in Table 3. In terms of strengths of the proposed learning program, almost all of the participants reported that the online learning program provided them with plenty of flexibility and convenience in learning. Most found the learning system easy to use and supportive for self-directed learning and practice with the support of expert feedback. Many students enjoyed building a mental map to represent their knowledge underlying the diagnostic tasks, which made their knowledge more solid and systematic. Accordingly, students commented that the learning program was helpful for improving their subject knowledge and diagnostic problem-solving capability. The participants also mentioned weaknesses of the learning program and provided relevant suggestions for improvement. First, they mentioned that more learning materials, particularly multimedia learning content (e.g., videos for clinical examinations) and diversified exercises (e.g., short questions) for self-practice could be added to the system. Second, students suggested including more learning cases into the online learning system for self-directed learning and practice. Third, they reported that some operations of the system (e.g., the diagram tool building mental maps) were complicated and could be simplified.

<table>
<thead>
<tr>
<th>Students’ comments</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strengths of the learning program</strong></td>
<td></td>
</tr>
<tr>
<td>Plenty of flexibility in learning, without time and space constraints.</td>
<td>24</td>
</tr>
<tr>
<td>The learning system is easy to use.</td>
<td>16</td>
</tr>
<tr>
<td>Expert support is helpful for self-directed learning and practice.</td>
<td>13</td>
</tr>
<tr>
<td>Constructing the knowledge in a mental map is helpful for thinking and learning.</td>
<td>12</td>
</tr>
<tr>
<td>Improving my subject knowledge, making my understanding more solid and systematic.</td>
<td>11</td>
</tr>
<tr>
<td>Enhancing my diagnostic problem-solving capability.</td>
<td>10</td>
</tr>
<tr>
<td><strong>Weaknesses of the learning program</strong></td>
<td></td>
</tr>
<tr>
<td>Some functions of the system are complicated, e.g., the diagram tool for building mental maps</td>
<td>7</td>
</tr>
<tr>
<td>Inadequate learning cases.</td>
<td>4</td>
</tr>
<tr>
<td>Inadequate learning materials.</td>
<td>3</td>
</tr>
<tr>
<td><strong>Suggestion for improving the learning program</strong></td>
<td></td>
</tr>
<tr>
<td>Inclusion of more diversified learning materials (e.g., videos for clinical examinations, short questions for self-practice).</td>
<td>7</td>
</tr>
<tr>
<td>Inclusion of more learning cases.</td>
<td>4</td>
</tr>
</tbody>
</table>
Discussion

Design of the model-based learning environment

The proposed GDeL environment is characterized by: (1) an authentic problem-solving experience in a simulated environment, (2) a complex process made accessible to students, and (3) the provision of expert support. The design is supported by the model-based learning approach: students are free to capture and reflect on their own models or task processes; meanwhile, the expert model is externalized to support and guide students to develop expert performance. The proposed design is aligned with the cognitive apprenticeship model. The learning environment involves an exploratory problem context for exploration with authentic problems. The complex process of diagnostic problem solving is scaffolded by highlighting its key actions involving clinical examination, intermediate judgment, and diagnostic conclusion. Learners are encouraged to capture and reflect on their problem-solving process during the task. Coaching is provided to individuals via feedback on their performance, which is associated with the expert model that helps learners to determine the difference between their performance and expert performance. Finally, based on the experience with multiple cases, learners are encouraged to articulate the knowledge underlying the problem-solving process.

Effects on learner perceptions and motivation for learning

Learners felt that the proposed GDeL environment was useful and easy to use. As a result, they reported a clear intention and motivation to use this approach. The participants also had positive perceptions of the cognitive strategies supported by the learning environment. Moreover, learners’ perceptions of the learning environment were significantly correlated with their perceptions of the cognitive strategies supported by the learning environment. Similar findings have indicated that instructional design is an essential factor that can influence learner satisfaction and acceptance of e-learning environments (Swan, 2001; Wang et al., 2013). Interview results supported the above findings. Learners found the GDeL system helpful in improving their subject knowledge and diagnostic problem-solving capability, and enjoyed the self-directed learning afforded by the expert support from the system and the online learning mode. In addition, students enjoyed using the diagram tool to build a mental map to represent their knowledge underlying their diagnostic processes. They found the mental mapping activity helpful for improving their thinking and systematic understanding. Similar findings have demonstrated that student learning outcomes in problem-solving contexts can be significantly improved when learners are able to externalize the complex problem-solving process and the knowledge underlying the task process with the support of computer-based cognitive tools (Wang et al., 2013; Wu et al., 2016).

Effects on knowledge achievement and problem-solving performance

After completing the learning program, the students had significantly improved subject knowledge and problem-solving performance in selecting clinical examinations to collect further information, making intermediate judgments based on examination results, and drawing diagnostic conclusions. The results are consistent with the perceived learning gains in subject knowledge and diagnostic problem-solving capability the students reported in interviews. Students’ achievements also echoed their comments on the benefits of expert support for the learning system and the advantage of constructing the mental map to articulate their underlying knowledge on the diagnostic tasks. The findings demonstrate the effects of model-based learning, which can be attributed to the representation of mental models and cognitive processes as well as the use of expert models to improve student performance. The findings align with the results of related studies (e.g., Ifenthaler, 2009; Schwarz, 2009).

Conclusion

Effective learning through problem solving is difficult to realize because learning in such contexts involves complex processes. Many learners tend to engage in surface rather than deep learning towards an in-depth understanding of the problem. Deep learning is characterized by a high level of engagement in learning driven by intrinsic motivation and more importantly supported by effective learning approaches or strategies that allow learners to manage complexity and key challenges (most on cognitive aspects) to sustain engagement and achieve a high level of understanding and performance (Wang, Kirschner, & Bridges, 2016). Owing to a lack of studies on how deep learning can be adequately empowered in complex problem contexts, the design and implementation of problem-oriented learning has been considerably dependent on the instructor’s personal
experience (Henry, Tawfik, Jonassen, Winholtz, & Khanna, 2012). Recent research has highlighted and demonstrated the importance of making thinking visible in complex situations, with a focus on self-constructed mental models or processes for deep learning and self-reflection by learners. With limited problem-solving capability and practical experience, it is still difficult for novices to develop expert-like performance without necessary support.

This study explores the design and effects of a model-based learning approach implemented in a web-based learning environment that allows learners to capture and reflect on their problem-solving process in visual formats as well as identify the gap between their performance and that of the expert when working with a number of clinical cases. The proposed approach attempts to utilize expert knowledge to transform open-ended problem-solving experience into a systematic and deliberate effort towards expertise development. The results show that the approach has promising positive effects on students’ achievements in subject knowledge and problem-solving performance and their perceptions and motivation for learning using the proposed learning environment. The effects of the proposed will be further examined by a control group design in a future study.

**Acknowledgements**

This research is supported by the General Research Fund (Project No. 17201415) from the Hong Kong SAR Government, and the Seeding Fund for Basic Research (Project No. 201611159071) from the University of Hong Kong. The authors would thank Professor Haijing Jiang for his valuable support for this project.

**References**


Deep and Surface Processing of Instructor’s Feedback in an Online Course

Kun Huang1, Xun Ge2 and Victor Law3

1Department of Instructional Systems & Workforce Development, Mississippi State University, MS, USA // 2Department of Educational Psychology, The University of Oklahoma, OK, USA // 3Program of Organization, Information, & Learning Sciences, University of New Mexico, NM, USA // khuang@colled.msstate.edu // xge@ou.edu // vlaw@unm.edu

*Corresponding author

ABSTRACT

This study investigated the characteristics of deep and surface approaches to learning in online students’ responses to instructor’s qualitative feedback given to a multi-stage, ill-structured design project. Further, the study examined the relationships between approaches to learning and two learner characteristics: epistemic beliefs (EB) and need for closure (NFC). Four emerging themes were identified where the students’ approaches to learning spread along a spectrum of deep to surface learning: number of feedback items addressed, understanding of feedback, quality in addressing feedback, and holistic thinking. In addition, the maturity of EB was likely to be associated with students’ understanding of feedback and the systematic and relational thinking demonstrated in their responses to feedback. The relationship was unclear between NFC and deep/surface learning characteristics. The findings provide implications for the design of feedback to scaffold deep learning in ill-structured problem solving.

Keywords

Approaches to learning, Deep learning, Epistemic beliefs, Feedback, Need for closure, Problem solving

Introduction

Providing feedback to students is an important instructional means in online learning. To facilitate online learning through feedback, we need to first understand how students process feedback. Using the theoretical lens of approaches to learning (Marton & Säljö, 1976a), this study examined online students’ deep and surface approaches to address instructor’s feedback, which was a complex, ill-structured task (Jonassen, 1997). Further, the study explored the possible relationships between students’ processing of feedback and two learner characteristics that might influence their approaches to feedback – epistemic beliefs (EB; Hofer & Pintrich, 1997) and need for closure (NFC; Kruglanski, 1990). We hoped that the findings would help researchers and online educators to recognize patterns of surface and deep processing of feedback, thereby helping learners of different EB and NFC to move towards deep learning.

Literature review

In this section, we first present theories related to approaches to learning. Next, we introduce ill-structured problem solving as a unique learning context in which approaches to learning need a greater depth of understanding. Subsequently, we review EB and NFC, which are likely to play a role in individual learners’ approaches to learning.

Approaches to learning

Ever since the seminal work by Marton and Säljö (1976a), approaches to learning have been characterized as a surface-to-deep continuum (Biggs, 1993; Marton & Säljö, 1976a). Departing from a mere cognitive account, Biggs (1987) conceptualized approaches to learning as both motives and strategies. At the deep level, learners are driven by an intrinsic interest in the subject matter of the task (Biggs, 1993). Accordingly, they adopt strategies to seek meaning and maximize understanding, including holistic (e.g., relating ideas; looking for principles) and serialist strategies (e.g., using evidence; examining the logic of argument; monitoring one’s own understanding) (Entwistle, 2000). Comparatively, surface learning is characterized by a motive to cope with the task, i.e., to meet requirements minimally, with reproductive strategies such as rote learning and concentration on procedures and isolated details (Biggs, 1987; Marton, 1983). Bigg (1987) accounted for learning with 3Ps: Presage, Process, and Product. While approaches to learning are driven by an intrinsic interest in the subject matter of the task (Biggs, 1993), they are affected by Presage factors, such as learners’ prior knowledge and personality. Together, Presage and Process lead to Product, which refers to learning performance.

ISSN 1436-4522 (online) and 1176-3647 (print). This article of the Journal of Educational Technology & Society is available under Creative Commons CC-BY-ND-NC 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Journal Editors at ets-editors@ifets.info.
Upon the development of quantitative instruments (e.g., Biggs, 1987; Entwistle & Ramsden, 1983), researchers have investigated approaches to learning in terms of their impact on achievement, effects of interventions, interactions with other constructs, or factors leading to deep or surface learning (e.g., García, Rodríguez, Betts, Areces, & González-Castro, 2016; Greene & Miller, 1996; Vos, der Meijden, & Denessen, 2011). Yet, research often yielded ambiguous or even contradictory findings regarding the effects of instructional strategies for deep learning (e.g., Dolmans, Loyens, Marcq, & Gijbels, 2016; Struyven, Dochy, Janssens, & Gielen, 2006). A few reasons have been given: first, while context-specificity is an assumption of approaches to learning, few studies examined learning approaches within a specific task; second, learners of different characteristics may exhibit different patterns in approaches to learning (Asikainen & Gijbels, 2017; Dinsmore & Alexander, 2012; Dolmans et al., 2016). Consequently, Dinsmore and Alexander (2012) called for the consideration of who, where, when, and what in investigating approaches to learning.

Adopting Biggs’ (1987) 3P model, this study responds to Dinsmore and Alexander’s (2012) call by examining approaches to learning in a particular learning context – ill-structured problem solving, and by investigating the roles of two learner characteristics or presage factors – EB and NFC (Figure 1). While the initial research on approaches to learning was conducted in the context of text recall or comprehension (e.g., Marton, 1983; Marton & Säljö, 1976a; Marton & Säljö, 1976b), little is known about the approaches learners undertake in solving ill-structured problems. As a complex process that involves qualitatively distinct yet interrelated stages, ill-structured problem solving may require learners to employ additional strategies that have not been accounted for in previous studies. The next section provides a review of ill-structured problem solving.

Ill-structured problem solving

Ill-structured problems are complex and ill-defined (Jonassen, 1997; Sinnott, 1989; Voss & Post, 1988). They do not have clear goals or definitive information. There may be multiple solutions, multiple paths to solutions, or no solutions at all. Ill-structured problems require problem solvers to go through iterative cycles between two essential stages (problem representations and problem solutions) while constructing arguments during the process (Ge & Land, 2003; Jonassen, 1997). Problem representation refers to defining features of a problem situation, identifying goals and subgoals, and figuring out a path from the initial state (i.e., initial situation of the problem) to the goal state (i.e., the desired condition, or the ultimate goal, to be reached through solution-searching activities) in the problem space. A problem space consists of various problem states, while a state is a representation of the problem in some degree of solution (Anderson, 2015). Due to the complexity of ill-structured problems, learners have to navigate across various problem spaces in order to identify the most relevant space (Sinnott, 1989). Frequently, the solver needs to go through an iterative process of redefining and re-representing the problem until an optimal solution is identified (Ge, Law, & Huang, 2016).

This study operationalized ill-structured problem solving in dual layers: (1) students worked on a semester-long instructional design project, which was an ill-structured problem itself; (2) students responded to instructor’s qualitative feedback to their projects, which was another layer of ill-structured problem solving.

![Figure 1. The key framework under investigation based on Bigg’s (1987) 3P model](image_link)
Epistemic beliefs

EB refers to individuals’ conceptions of the nature of knowledge and knowing (Hofer & Pintrich, 1997). An individual who holds less mature EB may believe, for example, that knowledge is certain and principles in a field do not change over time; on the other hand, an epistemologically more mature learner may question the authority of a textbook. EB has received an increasing attention in recent years due to its recognized importance in cognitive engagement, learning strategies, and achievement (e.g., DeBacker & Crowson, 2006; Muis, 2007; Schommer, Crouse, & Rhodes, 1992). It is reasonable to postulate that learners’ EB may have certain relationship with their approaches to learning manifested in the process of ill-structured problem solving, or in addressing instructor’s feedback as investigated in this particular study. Indeed, several authors have suggested such propositions (Chin & Brown, 2000; Entwistle, 2000; Marton & Säljö, 1976b), yet empirical research is needed to investigate the relationship between EB and learning approaches.

Need for closure

NFC refers to the desire for a firm answer to a question to avoid any forms of ambiguity (Kruglanski, 1990). The desire to reach closure is manifested in (1) seizing - arriving at a quick solution or using superficial processing, and (2) freezing - protecting prior knowledge or established solutions with only superficial scrutiny of new information (Kruglanski & Webster, 1996). NFC was found to be associated with learners’ cognitive engagement (DeBacker & Crowson, 2006; Harlow, DeBacker, & Crowson, 2011). DeBacker and Crowson (2006) projected that high NFC “compels students to accept the first reasonable solution they encounter, in order to end the ambiguity of problem solving” (p. 539). Because of the ambiguous nature of ill-structured problem solving, it is reasonable to anticipate that NFC may play a role in learners’ approaches to an ill-structured task. In this study, the task was to address instructor’s feedback to project.

Research questions

This study investigated approaches to learning in the process of ill-structured problem solving, operationalized in learners’ handling of feedback to their projects. Two research questions were asked:

- What characterize deep and surface approaches in students’ responses to instructor’s qualitative feedback to their problem-based projects?
- What are possible links between characteristics of deep/surface approaches and learners’ EB and NFC?

Methods

Both quantitative and qualitative methods were adopted in this study. Specifically, quantitative questionnaires served to identify learners’ EB and NFC profiles, and qualitative data were collected and analyzed to understand learners’ approaches to learning and their relationships with EB and NFC.

Participants, context, and instruments

Forty-four students were recruited from two sections of a completely online instructional technology course at a southeastern university in the United States. A major assignment of the course was a semester-long instructional design project that consisted of four progressive milestones leading to the final product – a training website for a target group of learners (Table 1). For each milestone, the instructor provided feedback using a set of grading criteria. In addition to assigning points to each criterion, the instructor provided qualitative comments (Figure 2). While most comments were on the criteria where students needed improvement, occasionally the instructor also provided “bonus” comments on future project work or on criteria in which students received full points. The comments usually did not give direct instructions, but prompted students to consider ways to improve their work. Students were required to respond to the feedback individually before working on the next milestone in the subsequent week. In their responses, students needed to (1) summarize the feedback, and (2) describe plans to address the feedback.
Table 1. Four project milestones leading to the final product (training website)

<table>
<thead>
<tr>
<th>Milestones</th>
<th>Schedule</th>
<th>Response to feedback</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Work submission</td>
<td>Week 6</td>
<td>Week 7</td>
</tr>
</tbody>
</table>
| 1          | Week 5    | Week 6 | • Identify target learners & training topics  
                              • Justify importance of the training  
                              • State general training goals  
                              • Identify applicable standards  
                              • Describe “big ideas” taught in the training  
                              • Identify learning objectives (target knowledge and skills)  
                              • Describe potential learning challenges  
                              • Start developing website (the About page) 2  |
| 2          | Week 7    | Week 8 | 3  |
| 3          | Week 9    | Week 11 | • Design an engaging Home page, with website introduction  
                              • Develop the Home page  
                              • Design instructional content supported by digital tools  
                              • Develop instructional content page(s) of the website 3  |
| 4          | Week 13   | Week 14 | 4  |

Feedback to Milestone 1 (Project Proposal)

Team 7

<table>
<thead>
<tr>
<th>Components</th>
<th>Points Received</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Team number and members’ names (2 pts)</td>
<td>2</td>
<td>I am very impressed that you’ve had two synchronous meetings.</td>
</tr>
<tr>
<td>2. Synchronous online meeting (3 pts)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3. Target learners (10 pts)</td>
<td>8</td>
<td>You have narrowed down the target learners, which is great. But what are some details or characteristics of the learners that may affect how you design the website for them?</td>
</tr>
<tr>
<td>4. Target teaching topic (10 pts)</td>
<td>10</td>
<td>Just to be sure, the main topic is “proper use and etiquette on social media website,” while “use online tools to research, discuss and collaborate” describes the learning activities to study the topic, correct?</td>
</tr>
<tr>
<td>5. Importance of the target topic (5 pts)</td>
<td>5</td>
<td>You have made a good case for the importance of the topic.</td>
</tr>
<tr>
<td>6. Goal of the instruction (10 pts)</td>
<td>10</td>
<td>I like your idea of using a video to demonstrate knowledge gain.</td>
</tr>
<tr>
<td>7. Coherence and logic (10 pts)</td>
<td>10</td>
<td>All the parts are indeed coherent.</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>You have picked a worthwhile topic, and the learning activities/assessments reflect active learning. I look forward to seeing how you move forward with the project. Meanwhile, I would like to remind you that this project requires students to learn from the website without an instructor on the side. So consider how you can design the website to complement the F2F ICT 1 course, and how to support students to learn by themselves online.</td>
</tr>
</tbody>
</table>

Figure 2. An example of instructor’s feedback

At the end of the course, the students completed two surveys (Table 2): an 18-item Discipline-Focused Epistemic Beliefs Questionnaire (Hofer, 2000) and a 15-item NFC survey (Roets & Van Hiel, 2011). Students’ EB and NFC scores were calculated and ranked. Following the maximum variation principle (Creswell, 2007), we purposefully sampled only those whose EB and NFC were both ranked at the top or bottom 10, thereby retaining nine students distributed across the four EB-NFC quadrants. Figure 3 depicts the relative positions of all the 44 students in the quadrants while highlighting the nine selected students. While the High NFC-Naive EB and Low NFC-Mature EB quadrants had four and three students respectively, the other two quadrants had only one in each. This uneven distribution is consistent with the positive EB-NFC correlations reported in the literature (DeBacker & Crowson, 2006).
Table 2. Subscales, sample items, descriptive statistics, and Cronbach Alphas for the EB and NFC instruments

<table>
<thead>
<tr>
<th>Subscales</th>
<th>Example items</th>
<th>Mean</th>
<th>SD</th>
<th>Cronbach α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemic beliefs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certainty of knowledge</td>
<td>Principles in this field are unchanging</td>
<td>3.09</td>
<td>.40</td>
<td>.78</td>
</tr>
<tr>
<td>Source of knowledge</td>
<td>If you read something in a textbook for this subject, you can be sure it is true</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Justification of knowing</td>
<td>Correct answers in this field are more a matter of opinion than fact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attainability of truth</td>
<td>Experts in this field can ultimately get to the truth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need for closure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>I enjoy having a clear and structured mode of life</td>
<td>3.93</td>
<td>.79</td>
<td>.88</td>
</tr>
<tr>
<td>Predictability</td>
<td>I dislike unpredictable situations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decisiveness</td>
<td>When I have made a decision, I feel relieved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguity</td>
<td>I don’t like situations that are uncertain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close-mindedness</td>
<td>I dislike questions which could be answered in many different ways</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Relative positions of all the 44 students in the EB-NFC quadrants (stars represent the nine sampled students)

Data sources and analysis

In addition to the EB and NFC data, there were three data sources: (1) instructor’s feedback to the four milestones, together with students’ responses to feedback, (2) students’ end-of-semester reflections describing their perceptions of the instructor's feedback, and (3) students’ project work.

To articulate the subtleties of approaches to learning as a general construct manifested in the context of ill-structured problem solving, we followed a bottom-up approach in coding: while being guided by principles of the approaches to learning in the literature, we allowed constructs to emerge from data (Biggs, 1993). In the first stage of data analysis, two researchers coded the nine students’ responses to feedback and their reflections, without the knowledge of the students’ EB-NFC profiles. Open coding was conducted, followed by axial coding to characterize distinct themes emerged from the data (Strauss & Corbin, 1990). Once the main themes were identified, the researchers then conducted the second round of coding within each theme to identify sub-patterns emerged from the data. While the levels of processing exhibited in students’ responses ranged along a continuum, we identified those that showed salient deep or surface learning patterns, given the focus of this study. In the second stage, the students’ EB-NFC profiles together with their respective quadrants were revealed, which were subsequently cross-examined with the deep and surface learning patterns demonstrated from their data. A linkage was identified when the data showed a clear association between a deep/surface learning pattern and EB or NFC. Throughout the coding, the researchers went through an iterative process of discussing...
interpretations and inferences, until consensuses were reached. Students’ reflections and project work were used to triangulate and provide necessary background information.

Results

**Question 1: Characteristics of deep and surface approaches to addressing feedback**

Before reporting the findings, we provide an overview of the nine students’ approaches to feedback. Among the nine students, Mark was the “poster” student who demonstrated deep processing in almost all aspects in his responses. On the other end, Chris’ responses demonstrated mostly surface approaches. The other seven students exhibited a mixture of deep and surface approaches.

The analysis of the students’ responses to feedback revealed four themes where students’ levels of processing spread along a spectrum (Table 3). In this section, we present each theme and report the most salient patterns that characterized deep and surface learning within the theme. It should be noted that the unit of analysis in this section is not individual learners, but the aggregated behavioral and thinking patterns exhibited in their responses and reflections.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Deep processing</th>
<th>Surface processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of feedback items</td>
<td>• Addressing most feedback</td>
<td>• Addressing minimal or limited feedback</td>
</tr>
<tr>
<td>items addressed in responses</td>
<td>• Addressing “bonus” feedback that did not require responses</td>
<td>• Addressing only feedback that required responses</td>
</tr>
<tr>
<td>Understanding of feedback</td>
<td>• Accurate understanding; appropriate problem representations</td>
<td>• Misunderstanding or simplistic understanding; inappropriate problem representations</td>
</tr>
<tr>
<td></td>
<td>• Elaborations that showed internalization and enriched problem representations</td>
<td>• Mere (or no) restatement of feedback</td>
</tr>
<tr>
<td>Quality in addressing feedback</td>
<td>• Application of general feedback to specific situations, with elaborations</td>
<td>• “Will do”</td>
</tr>
<tr>
<td></td>
<td>• Integration of feedback into existing understanding or project work</td>
<td>• Explaining causes of an issue without suggesting solutions</td>
</tr>
<tr>
<td></td>
<td>• Actual work on solutions beyond requirements</td>
<td>• Addressing feedback to a limited degree</td>
</tr>
<tr>
<td>Holistic thinking</td>
<td>• Solutions driven by explicit problem representations</td>
<td>• Fixing the obvious while leaving deeper issues unaddressed</td>
</tr>
<tr>
<td></td>
<td>• Mindful navigations between two problem spaces (feedback and project),</td>
<td>• Solution-orientedness without a clear link to problem representations</td>
</tr>
<tr>
<td></td>
<td>leading to transformed problem representations</td>
<td>• Addressing feedback without holistic considerations of the project</td>
</tr>
</tbody>
</table>

**Theme 1: Number of feedback items addressed in responses**

Across the four milestones, each team received a total of 11-17 feedback items from the instructor. Deep learning was demonstrated by two patterns within this theme. The first pattern was an attempt to address most or all feedback. For example, Mark explicitly responded to all of the 11 feedback items. The second pattern was regarding the aforementioned “bonus” feedback. Although responses were not expected, students such as Megan responded to both of the two future suggestions given to her project.

Surface learning within this theme was the most evident in Chris’ responses. Among the 14 feedback items given to his project, Chris responded to only two. Regarding the “bonus” feedback, half of the students did not acknowledge or address them.
Theme 2: Understanding of feedback

The students varied in their understanding of feedback, which was shown in two patterns. The first is the extent to which students demonstrated accurate understanding of the feedback. The second is the extent to which students internalized the feedback.

At the deep-learning end, across the four milestones, Caleb, Mark, and Sandy showed an accurate understanding of all but one feedback item. Figure 4 illustrates an example of the instructor’s feedback to Mark’s team project Milestone 1, and Figure 5 shows Mark’s responses to the feedback. Deep learning was further shown in the second pattern, when students went beyond an implicit acknowledgement of feedback by demonstrating internalization. A representative case was Jay’s response to Milestone 1 feedback. A requirement of Milestone 1 was to describe the target topics each team planned to teach with their proposed training website. The proposal of Jay’s team indicated their misunderstanding by mixing target topics (social media etiquette) with intended learning activities (“Student will use online tools to research, discuss, and collaborate”). When the instructor pointed out this issue, Jay responded that the instructor “wants to be sure that we clearly set the learning goals (topics) separately from the means of teaching (activities).” In his own words, Jay demonstrated internalization of the feedback, which helped to construct an appropriate problem representation conducive to a solution.

<table>
<thead>
<tr>
<th>Components</th>
<th>Points Received</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Target teaching topic (10 pts)</td>
<td>8</td>
<td>You seem to have a good idea about what the teaching/learning activities will be like. Teaching topics, on the other hand, have a somewhat different meaning. It is the main concepts students learn from learning activities. So, what are the main topics you plan to teach?</td>
</tr>
<tr>
<td>2. Importance of the target topic (5 pts)</td>
<td>4</td>
<td>Learning how and why electricity works is certainly important, but why is there such a need for your students to learn? Is it because the topics are on the curriculum, or perhaps the students have difficulties to grasp some of the concepts? More elaboration will strengthen the argument.</td>
</tr>
</tbody>
</table>

Figure 4. Instructor’s main items of feedback to Mark’s team project Milestone 1

On the other end of the spectrum, regarding the first pattern (i.e., demonstrating accurate understanding), two students, Jay and Megan, demonstrated complete misunderstandings of a considerable amount of feedback. Slightly better along the spectrum, there were cases of superficial or simplistic understanding. For example, Sharon’s team received seven items of feedback to their Milestone 1. In her response, Sharon did not address any specific feedback, but attributed all the issues to her team’s not being able to select a “teach-able” topic. This understanding was further confirmed in her later response to Milestone 2 feedback, where she felt that her team had done a better job ever since to make their project more “teach-able.” As shown, instead of trying to understand feedback at a deeper level, Sharon resorted to constructing a simplistic problem representation, and continued to stay in this problem space. The narrow and mis-conceptualized problem space might have prevented her from navigating appropriate paths to solutions. Regarding the second pattern (i.e., internalizing feedback), most students did not show internalization but merely restated feedback or did not restate at all. In fact, Mark was the only exception who routinely rephrased instructor’s feedback in his own words to show his understanding (Figure 5).
Theme 3: Quality in addressing feedback

While Theme 2 is about understanding of feedback, this theme is about how well students actually addressed feedback. Deep learning was evident in three patterns. First, some students were able to apply the instructor’s general feedback to specifics of their project, elaborating with details, examples, or questions. An illustrative example was Caleb’s response to Milestone 4 feedback. When the instructor pointed out that the stated learning objectives were not aligned with the instructional content on his training website, Caleb responded, “After thoroughly reviewing (the instructional content) and comparing it to (the learning objectives) on my website, I can CLEARLY see the misunderstanding that was taking place on my part.” In what followed, Caleb identified three specific learning objectives that were not addressed in the instructional content and discussed how he would address the issue.

The second pattern of deep learning was demonstrated in the investment of effort to actually work on addressing feedback, which was beyond the requirement of discussing plans to address feedback. Mark and Megan routinely did extensive work in their responses. For example, in Milestone 3, Mark’s team added a real-life scenario to their homepage to draw learners’ attention and introduce them to the website, but the instructor pointed out that the scenario was too lengthy to keep learners’ attention. In addition to recognizing the issue, Mark rewrote the entire scenario in his response.

The third pattern, which emerged as the deepest level of processing, was demonstrated in the students’ integration of feedback into their existing understanding. The pattern was evident in Mark’s response to Milestone 2 feedback. Milestone 2 required students to review and identify from three given sets of standards the ones applicable to their project. Originally, Mark’s team identified standards from one of the three sets. The instructor suggested that another set of standards might be more relevant to Mark’s project (i.e., electricity concepts for elementary students). Mark responded, “we originally decided ... the (first set of) standards ...
because the project does show creativity and innovation and also contains critical thinking, problem solving, and decision making (which was what the first set of standards was about). On the other hand with the (first set of) standards primarily associated with digital learning, the project will be better suited to follow (the suggested set of standards). As shown, Mark did not take the feedback at the face value by simply adopting the suggested standards. Rather, he reflected on and justified the reasoning behind the original decision and discussed how both sets of standards were relevant to his project. Through reconstructing his initial mental model, Mark updated his problem representation, which went beyond the scope of the feedback to the project level.

Towards the other end of the spectrum, four patterns characterized surface learning. The first pattern was shown in many responses as some variations of a “will do” statement. For example, Milestone 3 feedback requested Rose’s team to revise their learning objectives. Rose responded, “After speaking with the rest of my group, we will be updating our (learning objectives) accordingly.” However, no specific revision plan was proposed.

The second pattern, which showed a somewhat deeper level than a mere “will do,” was a tendency to explain the cause of an issue without attempting to suggest solutions. In her responses to both Milestone 3 and 4 feedback, Sandy discussed the causes (i.e., division of labor among team members) that had led to the issues pointed out in instructor’s feedback (i.e., crowded home page or inconsistencies across different pages), yet there was no discussion on how she planned to resolve the issues.

In the third pattern, students did attempt to discuss solutions, but to a limited degree. For example, Milestone 1 feedback asked Caleb to give more considerations to his target learners’ characteristics. In his response, Caleb described only one aspect of his learners (seniors) by referring to them as “the generation that never grew up with technology.” This limited scope of addressing feedback was quite common among the students’ responses, which was not conducive to successful problem solving.

In the last pattern, the students attempted to fix obvious issues while leaving deeper ones unaddressed. For instance, in Milestone 2, Jay’s team identified two big ideas they expected learners to take away from their training. The feedback asked the team to rephrase the first big idea from the learners’ perspective, and justify why they believed the second one to be a big idea, which was in fact a learning activity. In his response, Jay rephrased the first big idea, yet left the deeper issue (i.e., justifying the second big idea) unaddressed. Sandy’s response to Milestone 1 feedback showed a different approach to fixing obvious issues. The instructor suggested her team to consider their target learners’ characteristics, such as prior knowledge, motivation, and interest. Sandy responded, “We were instructed to take into consideration our target group’s prior knowledge … we could eliminate this issue if we narrowed in on a more specific group of students.” As shown, instead of trying to understand the learners, Sandy opted for a shortcut solution by narrowing down target learners to “eliminate” the issue.

At this juncture, it is important to point out that Sandy’s quick-fix solution was likely to be rooted in her narrow understanding of the feedback. Instead of interpreting the feedback as a prompt for considerations of a key instructional design principle – thorough learner analysis to inform the design of training, Sandy’s understanding of feedback focused on isolated details – learners’ prior knowledge only. Consequently, her solution was to narrow down learners to eliminate the “prior knowledge issue.”

Sandy’s case revealed a new dimension, problem space, in conceptualizing approaches to learning in the context of ill-structured problem solving. When learners did not gain a clear understanding of feedback and operated in a different problem space, they might not be able to reach appropriate solutions despite exhibiting certain patterns of deep learning. Megan was such a case – she invested extensive efforts in addressing feedback through elaborations, examples, and actual work, yet her solutions or plans for solution were often judged unsuccessful due to her misunderstandings of the feedback.

**Theme 4: Holistic thinking**

As shown in Theme 3, the close connection between problem representation and solution is a manifestation of holistic thinking, that is, taking into consideration the relationships among different components while solving a problem. Emerging from the data, two salient patterns demonstrated holistic thinking, or lack thereof: solutions driven by explicit problem representations, and mindful navigations between feedback and project.

At the deep-learning level, students’ solutions were clearly driven by explicit problem representations. Mark was a representative case. In addition to establishing an accurate understanding of most feedback, Mark turned
problem representations into explicit goals for his solution, demonstrated in statements like, “To help students better understand why learning how electricity works is important … we should …” As shown in Figure 5, Mark’s solutions routinely used these goal-setting statements, which were built upon his understanding of the feedback. Consequently, his description of the feedback, solutions, and actual work were all logically and cohesively aligned.

The second relational pattern of deep learning was shown in students’ mindful navigations between two problem spaces, the feedback and the project, in their responses. When deciding whether to keep a visual element on his website in response to feedback, Mark reasoned that he might remove it “unless it greatly improves the project during final review.” It is clear that Mark went beyond the problem space of the feedback and entered into the problem space of the project, by anchoring his planned solution in the project. Mark’s end-of-semester reflection corroborated the finding, “When reviewing feedback I would make sure to … look back at what we had originally done.” An even deeper level of processing was shown in Mark’s response to Milestone 4 feedback, where the instructor suggested moving an instructional video from one page to another on their website, in order to make the website structure clearer. Mark responded that he would move the video as suggested, and then completely delete the original page that contained only the video, “so that learners have the opportunity to learn (from the video) about how a circuit works before reviewing the material (on the new page).” In this example, Mark considered the instructor’s general feedback in light of the project itself. Through mindful navigations between the problem spaces of the feedback and of the project, Mark was able to transform his problem representation from “moving the video to a new page” to a more holistic level, with the consideration of causal relationships of various instructional design components. Indeed, Mark stated in his end-of-semester reflection that the feedback “made us think more thoroughly and gave us the opportunity to give clarity to our work.”

The two aforementioned relational patterns were lacking in the case of surface learning. A frequent pattern in the students’ responses was solution-orientedness, that is, jumping quickly to solutions without a clear link to problem representations. Megan and Jay were two representative cases whose responses often leaped prematurely into lengthy solutions. For example, when the instructor asked Megan’s team to consider the characteristics of their target learners (14-16 years old), Megan’s response did not describe how she understood the feedback but discussed extensively how graphic content, animations, sound, etc. might benefit teaching her target learners. Without explicit problem representations to drive her solutions, Megan often appeared to be approaching feedback by fixating on the solution, and volunteering extensive, yet often irrelevant and incoherent, details.

The second surface learning pattern was shown in a considerable number of responses that focused mostly on addressing feedback without a holistic consideration of how related components, at both the feedback and project levels, might be affected by a solution.

Question 2: Links between deep/surface approaches and learners’ EB and NFC

Regarding the second research question, we first provide an overview of the four EB-NFC quadrants, highlighting salient deep and surface learning characteristics of each, and then connect the characteristics to EB and NFC respectively.

The Mature EB-High NFC quadrant had only one student, Caleb, who exhibited both deep and surface learning patterns. He addressed only half of the feedback. However, in the feedback he chose to address, Caleb did show a good understanding of the feedback and proposed reasonable solutions, although his responses often lacked elaborations and details. Caleb’s understanding of and plans to address feedback showed cohesiveness and logical connections in general.

The Naive EB-Low NFC quadrant also had only one student, Chris, who had the lowest scores in both NFC and EB maturity among the 44 students. Chris addressed only two out of 14 feedback items, the least among the nine students. Most of his responses lacked understanding of the feedback. Although the lengths of his responses were comparable to others, he appeared to be operating in his own problem space. While his plans to address feedback did show some relation with his own, albeit inaccurate, understanding of the feedback, his responses generally lacked logic and cohesiveness.

The Mature EB-Low NFC quadrant had three students, Mark, Rose, and Sandy. Among them, Mark was the “poster” student whose responses exhibited almost all of the deep learning characteristics. Despite variations, this group generally exhibited good understanding of the feedback, although not all of them chose to elaborate
their responses. Their holistic thinking was demonstrated in the alignment between their solutions and understanding of the feedback, although sometimes implicitly.

The Naive EB-High NFC quadrant had four students, Cathy, Jay, Megan and Sharon. Despite variations, this group showed misunderstandings or simplistic understandings of a considerable portion of feedback. Similar to Chris in the Naive EB-Low NFC quadrant, this group’s responses generally lacked logical connections and cohesiveness.

Relating the EB-NFC profiles of the four quadrants to their respective approaches to addressing feedback, we first examined the possible links between EB and approaches to learning. The two quadrants of mature EB and the other two of naive EB appeared to differ in the following patterns: (1) understanding of feedback, and (2) logical connections and cohesiveness shown in students’ understanding of feedback and plans to address it. Compared with those with naive EB, the students in the mature EB quadrants generally exhibited better understanding of feedback, and showed better alignment and coherence in their responses.

Regarding the possible links between NFC and approaches to feedback, we were unable to identify any clear patterns that could distinguish the two high-NFC quadrants from the low-NFC ones. The only plausible link was regarding students’ responses to the aforementioned “bonus” feedback. Among all the seven (out of nine) students who received such feedback, all of the three students in the two high-NFC quadrants responded, yet none of the four low-NFC students did so. The final reflection of Caleb, a high-NFC student, may provide some insight into this phenomenon, “Paying attention to feedback ahead of time will save you in the long run.”

Discussion and implications

Utilizing both qualitative and quantitative methods, this study sought to develop a context-based conceptualization of different learning approaches while taking into consideration of individual learner characteristics (Dinsmore & Alexander, 2012). The study extends our understanding of the approaches to learning in a number of ways. First, this study investigated deep and surface learning patterns in students’ responses to instructor’s feedback, which represented an important, ill-structured learning task with multiple layers of problem spaces. While the existing literature identifies general characteristics of deep and surface learning (Biggs, 1993; Entwistle, 2000), the current study showcases how learners’ deep and surface approaches manifest in the particular processes of addressing feedback. The contextualized themes and sub-patterns correspond to what was suggested in the deep or surface learning literature. For example, deep learning was characterized as an intention to extract meaning (Entwistle, 2000) and an intrinsic motive to actualize interest and competence (Biggs, 1993). In the current study, the attempt to address most feedback was a manifestation of an intrinsic motive to understand the status of the project to further improve it. Comparatively, surface learning is characterized in the literature as an attempt to meet requirements minimally (Biggs, 1987), as well as a conceptualization of a task or situation as unrelated bits of information (Entwistle, 2000). In the current study, these characteristics were manifested in students’ simple restatements of feedback without elaborating their understanding, and in their focus on addressing isolated details in feedback while leaving out deeper issues.

Further, this study revealed the complexities inherent within deep and surface learning in ill-structured problem solving. Consistent with Marton and Säljö’s (1976a) work, the study suggested that deep and surface approaches were not dichotomous. The two approaches did not exactly mirror each other in an opposite way; rather, they each had their own patterns that spread along a spectrum. An additional complexity was evident in the simultaneous exhibition of both deep and surface approaches by the same individual such as Caleb, who showed deep learning patterns such as accurate understanding of feedback, while in the meantime demonstrated surface processing in responding to only half of the feedback and in lacking elaborations in responses. Chin and Brown (2000) reported similar findings that learners who were labeled as surface approach might exhibit deep strategies. Different from Chin and Brown (2000) who argued that learners have a predominant learning mode, we found it often hard to label a learner as predominantly deep or surface in the ill-structured problem-solving context.

The third contribution of the study was the identification of problem space as an important additional dimension of deep and surface learning in the context of ill-structured problem solving. In this study, the students were supposed to learn instructional design through working on an instructional design project (a broader problem space) and responding to feedback (an immediate problem space). In the case of deep learning, the feedback not only prompted Mark to address immediate issues in the feedback, but also pushed him to explore the broader space of the design project. He had to navigate between the two problem spaces (the feedback and the project) to
construct meaning and reach new understanding of instructional design. However, many students were unable to expand their understanding from the immediate to the broader space despite the instructor’s scaffolding through feedback. Consequently, they continued to stay in a limited problem space (i.e., a shallow representation of the feedback). For those students, even though they exhibited deep learning approaches, such as elaboration and self-questioning (Chin & Brown, 2000), their approaches failed to lead to fruitful solutions because they were not solving the “right problem.” At an even more surface level, students chose to narrow down their problem space, which led to a quick and tangible solution to close the case. For instance, instead of trying to understand more about target learners, Sandy opted to narrow down the learner group to eliminate the “prior knowledge issue.” For this approach of narrowing down problem spaces, we coin the term degenerative thinking.

This study further revealed how learners of different EB and NFC might show different approaches in processing feedback. The students with mature EB who believed in the changing nature and personal construction of knowledge were more likely to invest effort in meaning making to establish a genuine understanding of feedback. Such beliefs might explain their more accurate understanding of feedback, which led to logical solutions. On the other hand, different from expected, NFC seemed to have played a less influential role in processing feedback, with its only linkage to high-NFC students’ attempt to address bonus feedback. This phenomenon might be explained by the seizing aspect of NFC (Kruglanski & Webster, 1996), that is, to handle and clear everything out of the way early in the process in the hope of a smooth closure. However, a controversial case seems to be Caleb, a high-NFC student — while responding to bonus feedback, he only addressed half of the other feedback. More empirical studies are needed to clarify the roles of NFC and EB in learners’ approaches to ill-structured problem-solving tasks.

This study offers valuable implications for instructional designers and online instructors. By identifying issues and weaknesses as learners progress along problem-based projects in online learning environments, we can develop effective tools and strategies for diagnosis, formative evaluation, and scaffolding. Online instructors may use students’ responses to feedback as a strategy to pinpoint exactly the learning issues (e.g., causes of misconceptions) and follow up with appropriate scaffolding. Researchers and instructional designers may work together and apply the findings of this study to develop robust instruments for measuring students’ learning-in-progress along the continuum of deep and surface learning. It follows that a self-regulated scaffolding system should be developed to intentionally target on known weak areas, prompt learners to reflect on their learning process, and engage them in deep learning activities, such as checking their understanding of feedback, examining their problem spaces, elaborating responses, integrating feedback holistically into the project, mindfully navigating between different problem spaces, and taking appropriate actions to move from inappropriate into appropriate problem spaces for productive and feasible solutions.

Moreover, this study helps us to understand motivational antecedents such as EB and NFC that underlie learners’ problem-solving processes, which allows us to anticipate and address individual learners’ surface approaches and foster deep learning with appropriate strategies. With further validation of empirical studies, we may use such motivational antecedents as EB and NFC to predict deep and surface learning so that appropriate strategies can be devised to help learners move towards deep learning as they gain competence in the problem-based learning environment.

Conclusions

Due to its qualitative nature and the small sample size, this study was unable to yield generalizable findings. Further, this study chose to focus on learners at the top or bottom of both EB and NFC, while leaving out those with moderate EB-NFC profiles. While recognizing the limitations, we believe that this study makes a meaningful and practical contribution to online learning. As increasing numbers of courses are offered online, it is highly important that we understand how students process feedback, which is an essential strategy for online instruction. In doing so, we are able to optimize the process in which the instructor provides constructive feedback and students respond effectively to feedback. Lastly, this study is unique in utilizing learners’ EB and NFC as lenses to examine their deep and surface learning approaches in problem solving. The results of this study have generated further questions for future research. More quantitative or mixed-method studies can be conducted to verify the relationships between EB, NFC, and approaches to learning. Future studies can also examine learners with moderate EB-NFC profiles. Finally, instruments can be developed to measure deep and surface learning in the context of ill-structured problem solving.
References


Investigating the Effects of Authentic Activities on Foreign Language Learning: A Design-based Research Approach

Ildeniz Ozverir¹*, Ulker Vancı Osam² and Jan Herrington³

¹Eastern Mediterranean University, English Preparatory School, Famagusta, North Cyprus // ²Eastern Mediterranean University, Faculty of Education, Famagusta, North Cyprus // ³Murdoch University, School of Education, Perth, Australia // ildeniz.ozverir@emu.edu.tr // ulker.osam@emu.edu.tr // j.herrington@murdoch.edu.au

¹Corresponding author

ABSTRACT

Achieving communicative competency in English classes has been a key goal in contexts where English is taught as a foreign language (EFL). During this process, however, integrating the difficulty and complexity of real life tasks into classroom teaching has often been disregarded. Lack of opportunities for authentic language use often results in learners’ gaining extensive knowledge about the target language (know what) while they are weak in using the language in a meaningful way (know how). Accordingly, while learners can talk about grammar rules, they usually fail to use these rules for real communicative purposes in unstructured genuine settings. The present study employed a design-based research approach to investigate the use of authentic activities in EFL classes. For this purpose, an e-learning environment was created based on initial design principles of authentic activities and implemented in three pre-university level EFL classes in North Cyprus in two research cycles. Data were collected through semi-structured interviews, work samples, and observations. In accordance with the findings and continuous literature review, 11 design principles were derived from the initial design principles for the EFL context in order to facilitate competency-based foreign language use.

Keywords

Authentic learning, Authentic tasks, Foreign language education, Task-based language learning, Design-based research

Introduction

The inquiry voiced by Allwright (1979) as “Are we teaching language (for communication)?” or “Are we teaching communication (via language)?” (p. 167) addresses the issues related to communicative competence (Hymes, 1972) in English as a foreign language (EFL) contexts, where learners may have minimal opportunities to authentic use of the target language. When this is coupled with form-focused language teaching methods (as opposed to meaning-focused ones) in class, education often results in learners gaining extensive knowledge about the target language by being equipped with know what (Brown, Collins, & Duguid, 1989) yet having minimal or no chance to gain know how skills (Bowles, 2011; Rebuschat & Williams, 2012; Stalnaker, 2012), a problem referred to as “inert knowledge” (Whitehead, 1929). Harmer (2007) emphasises the inadequate structure of form-focused language teaching methods and argues that this type of approach “stops students from getting the kind of natural input that will help them acquire language because it fails to give them opportunities to activate their language knowledge” (p. 49). In such cases learners can talk about grammar rules, while they usually fail to use their knowledge for real communicative purposes.

As a remedy for the deficiencies caused by the form-focused language teaching methods, meaning-focused methods, specifically task-based language teaching (TBLT), have been proposed to give learners communicative competency so that they use the target language for communication (Ellis, 2003; Nunan, 2004; Willis & Willis, 2007). Ellis (2009), as a reaction to those who see TBLT simply as an exercise to do rather than a task to complete, points out that in TBLT learners address real problems and relate what they learn to everyday life. Likewise, constructivist learning environments support question-based, issue-based, case-based, project-based, problem-based or task-based learning, all of which differ in depth of complexity but are built on “... the same assumption about active, constructive, and authentic learning” (Jonassen, 1999, p. 219). Any task aimed at giving learners usable knowledge should parallel the real world as closely as possible, and the tasks that have real world relevance are “authentic activities” (Herrington et al., 2003). Authentic learning, according to Herrington, Reeves, and Oliver (2010), can situate learning tasks in contexts that close the gap between classrooms and real life. Based on this conception, this paper reports a study that provided learners with opportunities to use the target language in purposeful and complex ways through authentic activities.
Literature review

In order to inform the investigation, a literature review was undertaken across relevant areas of study including the “assistance dilemma” and authentic learning, as discussed below.

Assistance dilemma

Controversy about how instructional guidance in education should be provided has engaged many educational researchers (e.g., Hwang & Wang, 2016). While some researchers argue that explicit and direct instruction is the optimum method of education (Kirschner, Sweller, & Clark, 2006; Rosenshine, 2009), others argue that learners should be given opportunities to construct knowledge themselves under the guidance of more capable persons (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Lee & Anderson, 2013). This controversy has been referred to as the “assistance dilemma” by Koedinger and Aleven (2007). The following section will give information about direct instructional guidance and its drawbacks, and constructivist teaching methods.

Direct instructional guidance in language learning

According to this method, there is one correct answer to the problem and both the answer and the path are known by the teacher who is assumed to provide systematic detailed instructions to learners to achieve the outcome and the learning objectives (Kirschner et al., 2006). Supporters of this “traditional” method (Nunan, 2004) argue that syllabuses should be designed to teach from simple to complex structures “in building-block fashion” (Long, 1991, p. 41). This type of teaching employs “focus-on-forms” (Long, 1991) (or form-focused activities) aiming at teaching any one aspect of linguistic form at a time (“form of the day”) through explicit instruction (Ellis, 2012) in the hope that the language learner will, with practice, use the form correctly and appropriately in genuine settings.

Yet, such approaches were found limited in bringing the learner to the desired competence level to communicate effectively in the language (Ellis, 1997; Lightbown, 1985). Nunan’s (1995) research, for example, concluded that instruction on its own does not lead to acquisition and “the gap between teaching and learning will be narrowed when learners are given a more active role in three key domains of content, process, and language” (p. 154). Consequently, learning activities should be designed to be learner-centred, meaning-focused and communicative with real world relevance, as discussed below.

Constructivist teaching methods in language learning

Educators with more constructivist views advocate that “knowledge is not a thing that can be simply transmitted from one person to another” (Chee, 1995, p. 135). Johnston and Goettsch (2000), in the same vein, mention the difference between understanding (knowing what) and production (knowing how) in language education, as articulated by one of their study participants:

> They oftentimes don’t understand the rules. They just read a rule and go, “OK, I’ve read this since I was eleven years old. I have read it a million times back in my country and here.” And they’re still not using it right. They all know they need to use the third person singular “s” but half the class still doesn’t use it. They use it in the grammar exercises, but they don’t apply it while they are speaking or writing. (p. 456)

Unlike form-focused approaches – e.g., Grammar Translation Method or Presentation-Practice-Production (PPP), meaning-focused approaches with task-based, problem-based, and project-based activities shift the focus from the “forms” to “task completion” in which meaning is primary (Nunan, 2004; Willis & Willis, 2007). Linguistic elements (“form”) are still important and brought to learners’ attention as they arise incidentally in lessons and as a part of communicative language practice (“focus-on-form,” Doughty & Williams, 1998). In this research, form-focused instructional guidance is considered as direct teaching method whereas meaning-focused approaches are considered as minimally guided constructivist approaches. Table 1 summarises form-focused and meaning-focused methods of instruction in TESOL (Teaching English to Speakers of Other Languages).

Meaning-focused method is also recommended by Council of Europe (2001) when designing a curriculum. Tasks rather than linguistic structures should be the core units that describe the selection of goals. In other words, the objectives of the lessons should not be describing what specific language features students gain, like in the example: “The students will learn how to use the past continuous and past simple tenses to express an
interrupted action” (Ritchie, 2003, p. 114). Instead, the objectives should focus on using the language for an authentic purpose to be able to function in society, as is described in The Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001): “[Learner] can describe plans and arrangements, habits and routines, past activities and personal experiences.” In short, as Van den Branden (2006) suggests, “there should be a close link between the tasks performed by learners in the language classroom and in the outside world” (p. 6), which is authentic learning. To put it differently, meaning-focused method and authentic learning overlap in terms of relevance, meaning, purpose and type of the instruction.

Table 1. Instructional models of linguistic forms in TESOL

<table>
<thead>
<tr>
<th>Form-focused methods</th>
<th>Meaning-focused methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proactive</strong>: The form of the day is planned by the teacher to ensure that form is learned or practiced. If the form is not learned or practiced, the learning activity is not considered successful.</td>
<td><strong>Reactive or corrective feedback</strong>: Possible forms may or may not have been determined prior to the activity. Learners’ errors that arise while completing the activity define what forms will be focused on to be able to complete the activity.</td>
</tr>
<tr>
<td><strong>Planned</strong>: The form of the day is determined prior to the activity; however, learners are not explicitly made aware that a specific form is being studied and thus they act as language users rather than language learners.</td>
<td><strong>Incidental</strong>: Forms are not determined prior to the activity. They arise as learners complete the activity. Thus, instead of recycling a single form several times, a variety of forms are addressed based on the demands of the learners to be able to complete the activity.</td>
</tr>
</tbody>
</table>

In order to gain a deeper understanding of how meaning-focused methods help the learners in class, a number of researchers (Granena, 2016; Namazian Dost, Bohloulzadeh & Pazhakh, 2017; Nazari & Tabatabaei, 2016) compared traditional teaching with meaning-focused ones, more specifically TBLT, concluding that minimal guided methods (i.e., meaning-focused ones) outperformed guided (i.e., traditional) methods. Yet, these studies can be critiqued as the tasks they employed were mostly school type of tasks such as “spot-the-difference,” “find a suitable title for the original text” (Granena, 2016) or “sequence pictures” (Nazari & Tabatabaei, 2016), rather than real-world tasks. Although some studies (Guariento & Morley, 2001; Mishan & Strunz, 2003; Nazari & Tabatabaei, 2016) provided guiding principles to design TBLT, elements to assist designing real-world relevant tasks were missing. This study, then, seeks to provide guiding principles for teachers to design tasks which go beyond exercises in order to close the gap between the classroom and the real world by investigating the effects of authentic activities on foreign language learning. The next section will describe authentic learning and the characteristics of authentic activities.

**Authentic learning**

Authentic learning is an instructional approach that provides learners with opportunities to develop knowledge “embedded in the social and physical context within which it will be used” (Herrington et al., 2010, p. 15). Learners are given situations based on a major relevant real-world task which requires them to investigate a problem in depth in collaboration with peers and suggest their solution in the form of a product to be shared with a wider audience. In this respect, Herrington et al. (2003) describe authentic activities as those that:

- have real world relevance
- are ill-defined, requiring students to define the tasks and sub-tasks needed to complete the activity
- comprise complex tasks to be investigated by students over a sustained period of time
- provide the opportunity for students to examine the task from different perspectives, using a variety of resources
- provide opportunity to collaborate
- provide the opportunity to reflect
- can be integrated and applied across different subject areas and lead beyond domain specific outcomes
- are seamlessly integrated with assessment
- yield polished products valuable in their own right rather than as preparation for something else
- allow competing solutions and diversity of outcome.

The learning activity designed in the present study was initially guided by these ten characteristics, aiming to enable learners to gain robust knowledge that they can transfer to real-life. With this aim, the study investigated the following research question: In what ways do students achieve foreign language competency through the use of authentic activities? The following section describes the methodology used to guide this research.
Method

A design-based research (DBR) study was conducted, comprising four phases (Reeves, 2006) in two full iterative cycles of enquiry, each cycle lasting six weeks (Figure 1).

**Figure 1. Four phases of design-based research (Reeves, 2006, p. 59)**

The first phase of both cycles investigated the problem by consulting practitioners, using personal experiences, conducting a literature review and analysing reports on the causes and effects of the problem in practice. In the second phase, the learning environment was developed (as described below) based on the initial principles. This phase was followed by the first implementation—the third phase of DBR. It was a testing and refinement process so that the researchers fine-tuned the research by making changes to the learning environment, editing the design principles, and implementing the cycle again. For the refinement of the principles, coding was used as suggested by Miles and Huberman (1994). During the fourth phase, researchers reflected on their analysis of data, and shared their experiences and the outcome of their research in the form of 11 refined design principles (Ozverir et al., 2016) that can guide future educational practice. In this way, the findings contribute to both theory and practice.

Data was collected through documentary evidence from six students and two teachers from two different classes (three students from each class) during the first cycle, and from four students and one teacher during the second cycle. The participants were students studying English at the Preparatory School of an English-medium university in North Cyprus in order to follow their chosen subjects. The participants were heading towards B1—Threshold–level language competency (Council of Europe, 2001) and all the learning activities were designed to help learners achieve at this level. The activities were assessed based on the CEFR descriptors and Verification of learning approach (North, 2007). Three teacher participants were experienced staff members in the School.

The learning environment

The study was administered over a full semester. The learning activity was based on a scenario where the class was the editorial board of the City Newsletter, the teacher was the Editor and the learners were journalists. For the major task, learners were required to conduct research, individually or in pairs, to collect data on a problem that had a social significance, and write an article on it to be published in the newsletter, and present it orally. Students were given a list of problem topics or could choose their own and had to propose a possible solution. Subtasks such as producing posters and videos, were intended to inform the audience further. Students were also required to complete such scaffolding learning activities as analysing the structure of an article (see Figure 2 below) and giving effective presentations where they were expected to reflect on both their learning and the process.

Moodle (see https://moodle.org/) was used as the platform which enabled students to participate in asynchronous chats and access learning resources by providing file sharing functionality (Figure 3).

Assessment was based on the can do statements of CEFR. A link was provided to inform students of the assessment breakdown showing the different components and how they contribute to the total grade, making it transparent for learners.

The task cycle was predominantly concerned with meaning. Learners were encouraged to reflect their opinions and solutions throughout the task. However, at different stages in the task cycle the focus was also on language and how it functions in context (as it is recommended in characteristic 10 below). This was in parallel with
Willis and Willis’s (2007) claim that: “a focus on language occurs naturally when learners pause in their attempts to process language for meaning and switch to thinking about the language itself” (p. 113), enabling learners to develop their language skills independently of the teacher as they took their own initiative and produced “a far wider repertoire of language to express themselves” (p. 113). In the following section, findings related to learners’ written interaction (during online discussions) and production (i.e., the article they wrote) as well as their spoken production (presentation and artefact) are described in detail.

**Advantages and Disadvantages of Internet**

**Introduction**

Nowadays, Internet has an important part in our lives. Everyone is using the internet to get information about different topics as the internet has a lot of useful sources.

Internet offers us a lot of valuable resources. Firstly, we can visit a lot of websites and get information. For example, students can search for information in order to do their homework. Secondly, businessman can communicate with their clients, advertise their products, sell their products online. For example, you can get information about the latest mobile phone through the internet. Thirdly, everyone can use Internet for communication purposes. For instance, we can communicate with our loved ones when we are far away from them.

**Body**

Apart from the advantages, there are several dangers of using Internet. Firstly, there are some bad people out there. Therefore, we must be very careful when we chat online and we should not trust what people we really don’t know. We shouldn’t give our personal details, our address, our phone number to those we do not know well because they can steal our identity and do bad things. Secondly, we should not give our credit card details when we write emails. Also, we need to shop on safe sites such as Amazon, hepsiburada.com, or ebay and use our credit card details only on these safe sites in order to prevent online robbery and theft. Finally, some students are obsessed with Internet use and they spend hours and hours on the net instead of studying. As a result, they fail their exams. For instance, last year my friend was using the internet a lot and he only studied a little. At the end of the year, he failed his university exams.

**Conclusion**

In conclusion, using Internet is a great opportunity for all of us to improve ourselves while doing research as well as to communicate with each other. However, we need to be careful when we are online which is important for our safety. I strongly believe that Internet might not be as harmful as we might think of if we are careful and protect ourselves online.

*Figure 2. A screenshot of one of the learning resources on analysing the structure of an article*

*Figure 3. A screenshot of the homepage of the learning environment*
Results

Written interaction (online discussions)

Learners contributed to teacher-initiated online discussions. Their writing was evaluated in terms of relevance of content, stating a clear opinion, and contribution to the development of ideas. In this respect, the assessment process was guided by the descriptors given below:

Understanding and responding appropriately to the topic

Learners were expected to show competency by taking part in a number of online discussion forums. The aim was to provide learners with as many opportunities as possible to use the target language for meaningful communication and scaffold their learning process. When posts were analysed, it was observed that learners generally responded to questions appropriately, showing a clear understanding of what was asked. For example, when asked to examine the elements of an essay, they indicated the important elements according to the paragraphs, as one student responded:

I think the introduction paragraph must give general information about the article, not detail (A post by Seyit)
(N.B. All names are pseudonyms and quotes given are verbatim).

Expressing ideas and supporting them with examples

Online discussions provided opportunities for learners to use a variety of language structures. Learners used different structures to complete the subtasks. Some learners preferred to mention their opinion directly by writing “in my opinion…” or “I think….” Others preferred to provide their opinion implicitly, for example, by using a superlative:

Examples, example linkers, sequencing linkers, and cause/effect linkers have to be used in the body paragraph.
Linkers are the most important thing because the meaning of the paragraph can be followed easily. (Emre)

Generally, learners provided examples to support their opinions by referring to the places or information they were writing about.

Contribution for deeper understanding of discussions

Learners contributed to discussions for deeper understanding of issues relevant to their topics or, in some cases, for deeper understanding of the task requirement. Learners accomplished this by responding to the teacher-initiated discussions with initial comments and then responding to the teacher’s further comments. This was beneficial to learners because the teachers’ feedback guided them in focusing or re-focusing their attention. An example of this took place between Caner (the participant teacher) and his student Mert, on using a source. The guiding questions were: (i) What is a source? (ii) What is the role of a source? (iii) Do you think that a source is useful?

Mert provided his ideas:

If we have a source, we can easily find about what is the topic. It help us for give an information…I think, it's very important. When you do not know anything about your work. You cannot doing anything, therefore you have to get some help from it. (Mert)

Caner acknowledged Mert’s ideas and invited the other learners to contribute to this discussion:

Exactly, sources help us find information about a topic that we are researching. Without sources, we may not provide enough details or scientific evidence, etc…Thank you Mert…What do the others think? (Caner’s feedback to Mert’s post)

Mert continued his contributions to the discussion by explaining how he made use of sources. Then, Caner warned Mert of the dangers of relying too much on sources:

If you depend on the source completely, or if you use all the source, with the necessary and unnecessary parts, it is a lot of work, waste of time, but more importantly, stealing information by copy paste; this is called PLAGIARISM (information theft) and is against the academic rules...Many thanks Mert… (Caner)
From the above example, it is evident that a meaningful discussion took place between Caner and his student, Mert, regarding the use of sources and its benefits. Also, learners gained a deeper understanding of the task when the teacher took the opportunity to warn his students about plagiarism, an example of scaffolding. However, contributions to the online discussions were not free of weaknesses. Many cases were observed where learners ignored the teachers’ feedback and gave no further response.

**Showing understanding of others’ posts by commenting on their ideas**

In social situations, it is important that people show understanding of what other people say. In this respect, a more natural, colloquial talk in written form was observed and recorded on the discussion forum. This natural talk was first in the form of agreement/disagreement of their peers’ opinions by using basic structures such as “I agree with…” and then responding to the original question raised by the teacher:

I agree with … I think the body paragraph must tell us about main idea or topic and we must understand what is it about. (Burak)

**Asking for and giving clarification**

Asking for clarification when one is confused is a natural process of understanding. As such, the learning environment incorporated opportunities for learners to ask questions. However, learners did not seem to make use of this feature within the learning environment. Instead, learners approached teachers and verbally asked for help. Even though from the discussion above, it is evident that learners have displayed the ability to accomplish the objectives of the activity and its subtasks with the necessary competencies, learners seemed to lack awareness of the importance of responding to others’ posts or providing clarification when an issue had not been dealt with.

**Written production (article)**

CEFR “can do” statements, referring to written production, have driven the evaluation of the written artefacts in terms of content and linguistic competency. Each of these is described below in more detail.

**Content**

Content was analysed in terms of task fulfilment, coherence and unity of the written artefact. Task fulfilment of the articles was predetermined by the task instructions which were to gather information and bring solutions or suggestions for improvement to a problematic issue. Six descriptors that guided the assessment of the content of the compositions are described below.

Figure 4 shows an example of a student’s written product which was analysed to see whether the learning activity had provided the necessary conditions and opportunities for the learners to accomplish the tasks at the required level. The learner, Emre, chose the task of providing suggestions to the Mayor of Famagusta on entertainment for young people.

When Emre’s composition on entertainment is analysed with respect to the task, it can be seen that this student:

- can develop an argument with justification: He has provided reason why Famagusta does not have enough entertainment venues for young people by indicating the government’s lack of time. He also mentioned that the reason for choosing the water activity is that there is currently nothing similar to it available at present, illustrating that he:
- can support argument with relevant examples: In the composition, Emre has exemplified how he believes the activity is fun by describing it and thus demonstrates that he:
- can organise ideas in a logical manner: Emre’s composition was well paragraphed. He initially described his reason for choosing this topic and continued to describe the activity in detail. He concluded that he had no experience of the activity but predicted it will be popular among young people like him. Emre, thus, conveyed his ideas clearly.
This shows that the Examining Essays forum has supported learners to develop knowledge by applying compositional elements to produce a coherent and unified piece of writing. Emre has also shown evidence that he:

- can select appropriate information to address target audience: The information that Emre chose to use, such as the way the balloon functions in water, was appropriate, as he needed to indicate why he believed the activity would be fun and enjoyed by young people in order for the Mayor of Famagusta to finance it. Accordingly, he has also shown he:
- can locate desired information to support idea/s: Since Emre was able to describe the activity with information that he located from a source, it can be said that he has demonstrated the ability to locate desired information to support his ideas.

However, although some learners felt the need to use information from other sources, not all referenced these correctly. For example, in his composition Emre provided a link to indicate that he made use of a source to support his ideas. However, this lacks in-text referencing so it can be concluded that, he:
- can use information located from sources to support idea/s, but without reference: Referencing has been identified to be a weakness of the learning environment and thus it was planned to add instructional materials and activities for the next teaching cycle.

The overall analysis of the written product shows that learners selected appropriate information to address the intended audience, organised information logically, and therefore, developed an argument appropriate to their chosen task and justified it accordingly by providing examples to support the argument. This indicates that the learning environment assisted learners in developing their written production skills at the required level – B1.

Linguistic competency

The CEFR is a guiding document that does not prescribe, but describes how a learner functions as a social agent at the six levels (A1, A2, B1, B2, C1 and C2). The core of the CEFR is a set of communicative language activities and communicative language competences (North, 2007). The communicative competencies are subdivided into linguistic, sociolinguistic, and pragmatic competencies. For this research, focus was on linguistic
competencies. This is again subdivided into range (general linguistic range and vocabulary range) and control (grammatical accuracy and vocabulary control). The CEFR does not specify which linguistic features or lexical items are characteristic of each level but rather embeds them in “can do” statements that describe how the learner’s language should function for each level.

One of the propositions of this study was that authentic activities provide the opportunity to use the target language in context as it is used in real life, unlike traditional school type activities that provide limited opportunities to use a rich range of vocabulary and linguistic features. This proposition is expanded below.

Vocabulary

Lists of vocabulary developed for B1 level contained 2000 words (Milton, 2010). Thus, targeting the first 2000-word level provided the necessary opportunities for learners to learn and practice these words. Researchers argue for the fundamental importance of the first 2000 words and advise that:

If learners do not know the most frequent 2000 or 3000 words in English, they will have severe difficulties in understanding most written and spoken text and it will make it even more difficult to engage actively in written and spoken communication. (Stæhr, 2008, p. 150)

For this research, Paul Nation’s “Range program with GSL/AWL list” was used to analyse the range of vocabulary used by learners throughout the activity. During the data collection process, it was discovered that all learners had read all the posts on the discussion forums, together with all the articles written by their peers. Analysing the text and articles used on the discussion forums provided rich data on the vocabulary learners were exposed to in written form.

For each class, all text on the discussion forums, including the questions written by the teachers and all published articles, were analysed. Some adjustments were made; for example, proper names were removed. Table 2 displays vocabulary used by Ceyda’s (another participant teacher) class. Learners were able to practice 1745 different words in context, and 567 of these words were from different word families of the first 1000 most frequent words. Accordingly, learners were able to practice more than a total of 949 different words from the different family types. The other two classes had a similar outcome.

Table 2. The output created by Range program based on the vocabulary used in Ceyda’s class

<table>
<thead>
<tr>
<th>Word list</th>
<th>Tokens/%</th>
<th>Types/%</th>
<th>Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>First 1000</td>
<td>13025/82.86</td>
<td>983/56.33</td>
<td>567</td>
</tr>
<tr>
<td>Second 1000</td>
<td>1188/ 7.56</td>
<td>271/15.53</td>
<td>197</td>
</tr>
<tr>
<td>Third 1000</td>
<td>919/ 5.85</td>
<td>247/14.15</td>
<td>185</td>
</tr>
<tr>
<td>Not in the list</td>
<td>587/ 3.73</td>
<td>244/13.98</td>
<td>Unknown</td>
</tr>
<tr>
<td>Total</td>
<td>15719</td>
<td>1745</td>
<td>949</td>
</tr>
</tbody>
</table>

This finding confirms the previous research studies which claim that incidental learning through reading has an important place in language education (Waring & Nation, 2004) and learners in higher levels (as in B1) can learn new words by meeting the new word fewer times in comparison to learners at lower levels (Zahar, Cobb, & Spada, 2001).

Grammar

The CEFR suggests a grammatical span across the levels by describing the domain that defines the use of certain semantic functions. Table 3 below represents to what degree the learner can use the language and is tabulated as grammatical accuracy in B1.

Table 3. Grammatical accuracy in B1 (Council of Europe, 2001, p. 114)

<table>
<thead>
<tr>
<th>Grammatical accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express</td>
</tr>
<tr>
<td>Uses reasonably accurately a repertoire of frequently used ‘routines’ and patterns associated with more predictable situations.</td>
</tr>
</tbody>
</table>
This part of the study was designed to determine if the foreign language learning environment could provide opportunities for learners to develop their language skills in their journey towards becoming Independent Users of English at B1. Descriptors are expanded below.

Using a sufficient range of language to express ideas

In Emre’s composition, there were a sufficient range of complex sentences and grammar structures present to cater for the requirement of the B1 level. He used different grammatical structures appropriate to conveying his message, for example, conditionals and passives. Though these were not used frequently, he chose these structures to state what happens under certain conditions or when focusing on action. In general, Emre preferred to use simple structures, such as present simple when generalising, or giving factual information although this did not impede his overall communicative adequacy. He was also able to use language with good control for adequate communication. As Kuiken, Vedder, and Gilabert (2010) mention, communicative adequacy and to what extent the learner is able to complete the task is influenced by accuracy rather than the complexity of grammar.

Spoken production (presentation and artefact)

The broad CEFR “can do” statements referring to spoken production—“can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points”—has driven the evaluation of the presentations in the terms of content, presentation skills, and collaboration. Eight descriptors from the CEFR B1 level guided the overall assessment of spoken production: Learners can: (1) select appropriate information to address the target audience; (2) explain the main points relating to the topic with reasonable accuracy; (3) understand and answer most questions asked about topic; (4) ask questions to support further understanding; (5) maintain eye contact to hold attention; (6) speak clearly with little or no hesitation; (7) select appropriate visuals to support topic; (8) collaborate with partner and share work load.

Emre’s spoken production was typical of other students’. He was able to fulfil most of the “can do” statements. For example, he chose appropriate information to address the target audience, information as to why he thought his chosen type of entertainment would be best for the young people of Famagusta. He chose to describe the activity and also add an element of excitement by referring to challenges and fun, and also mentioned why he believed that water zorbs should come to Cyprus:

Maybe water zorbs can be like competition. As everyone knows students want to every time fun. … Cyprus is an island therefore this sport must be in Cyprus. I think water zorbs is suitable for Cyprus because Cyprus has too many students. (A script from Emre’s presentation)

His language was accurate enough to convey his message. Additionally, it was clear that the discussion on giving effective presentations on the discussion forum was useful since Emre, like all other participant students, was able to manage eye contact, use visuals to support his presentation and speak clearly with little hesitation.

Discussion

One of the key outcomes of DBR is a set of context-based design principles. Based on the findings and the continuous literature review, 11 design principles were derived from the initial design principles for authentic activities in EFL, and we argue that appropriate implementation of these characteristics have the potential to provide necessary conditions for language acquisition to occur in EFL contexts. Thus, authentic activities:

- have real world relevance
- are complex and ill-defined, requiring students to define the tasks and sub-tasks needed to complete the activity over a sustained period of time
- provide the opportunity for students to examine the task from different perspectives, using a variety of resources
- provide the opportunity to collaborate
- provide the opportunity to reflect
- lead beyond domain- and skill-specific outcomes
- are seamlessly integrated with assessment
- yield polished products valuable in their own right rather than as preparation for something else
are open-ended allowing competing solutions and diversity of outcome
are conducive to both learning and communicating
provide motivational factors

The use of authentic activities has led to the creation of a computer-assisted foreign language learning environment where learners interact with each other (see characteristic 4) on various topics over a sustained period of time (characteristic 2) through the discussion forum contributions (characteristics 3 and 5) for a real purpose (characteristics 6 and 9) that, in turn, assisted in the development or practice of various language skills in context (characteristic 1). It should be noted that students were enthusiastic and mainly completed the required tasks (characteristic 11). Learner motivation was achieved through providing ownership on the problem and the process to solve it, providing a challenge that is tied to a goal, providing freedom and control, providing authentic roles, targeting an authentic problem and authentic audience, and publishing student work (characteristics 6 and 10).

Written interaction (online discussions)

Contributions to the discussion forums indicated that learners developed their language skills under three broad sub-headings–content, opinion, and contribution to understanding. Their posts provided evidence that they understood what was expected and responded appropriately. The learners’ language was not always accurate but their message was understandable. Another attribute displayed was expressing and supporting opinion. Discussions supported knowledge development on issues such as the importance of using a source and the components of an essay. Learners displayed the ability to contribute to the discussion forums to support further understanding, though this was in most cases limited to their first contribution. Additionally, learners were exposed to an extensive amount of vocabulary and language as they gave feedback on others’ essays on the discussion form. They had the opportunity to use language for a real purpose, develop knowledge on different topics and improve language competency through the discussion board contributions and collaboration.

Written production (articles)

In the articles, learners developed an argument appropriate to their chosen task and justified it accordingly by providing examples, organizing ideas logically, and selecting appropriate information for the target audience. The two sub-headings (content and linguistic competency) drove the analysis of the articles. This indicated that the learners’ interaction with the learning environment through discussion forum contributions assisted both content and linguistic development on the articles.

Spoken production (presentation and artefact)

During the presentations, content, presentation skills and collaboration were examined. Learners were able to provide details of their topic and most selected appropriate information. Presentation skills indicated that the discussion forum on presentation tips was effective. During the presentations, learners were also required to reflect on their learning and the process. This class collaboration assisted learners in developing awareness that the activity provided them with opportunities to learn and practice know how, and supported each other in language development.

Limitations and future recommendations

Although the study has mostly achieved its aims, there were some limitations. First, the size of participants (10 students and 3 teachers) may seem small to generalize the results for larger groups. Nevertheless, each of these respondents provided a wide range of data (both written and spoken) which enabled the researchers to delve deeply into how the new learning environment increased the EFL students’ authentic learning opportunities. It is likely that if applied in larger groups similar results would emerge out of the study. Second limitation is related to the learning environment. In future iterations, more guidance to have learner-learner interaction for effective responses in giving and seeking clarification should be promoted. There is also a need to add guidance for referencing to avoid plagiarism. In terms of the spoken production of learners, the interaction between the presenter and his/her audience needs addressing, for example, by asking questions. Moreover, a third cycle is
suggested to assist teachers by providing training for a smoother implementation process. It may also be logical to replicate the study by redesigning tasks for lower CEFR levels such as A1 and A2.

Conclusion

The findings indicate that the learning activity was successfully linked with the characteristics of authentic activities to develop the relevant competencies at the B1 level. In particular, learners were able to express their ideas and opinions on various topics. They were exposed to an extensive range of vocabulary and language. They had multiple opportunities to use the target language in context. Higher order thinking skills also developed as learners were provided with opportunities to analyse and discuss others’ work and make judgments on how to improve it. Consequently, the findings suggest that the learning environment based on 11 characteristics of authentic activities in EFL promoted the use of the target language as a tool to convey messages adequately, both in written and spoken form, and facilitated learning new linguistic elements in context. Another key implication of authentic activities is that the learning environment lent itself to close the gap between language acquired in foreign language learning settings and the real world.

We strongly believe that in contexts where the target language is not spoken outside the school, using the guiding characteristics of authentic activities will provide opportunities for learners to use the target language in context for a purpose. This in turn will facilitate the internalization of the newly acquired linguistic structures. Moreover, teaching in language classes do not—and should not—focus exclusively on teaching a language but also on developing higher order thinking and problem-solving skills, also referred to as life skills. Thus, we propose that designing tasks taking into consideration the characteristics suggested in this study will provide learners with the opportunity to develop not only their language skills, but also important life skills.

References


Visualizing the Complex Process for Deep Learning with an Authentic Programming Project

Jun Peng¹, Minhong Wang²* and Demetrios Sampson²,³

¹KM&EL Lab, Faculty of Education, The University of Hong Kong, Hong Kong // ²School of Education, Curtin University, Australia // ³Department of Digital Systems, University of Piraeus, Greece

u3002116@connect.hku.hk // magwang@hku.hk // demetrios.sampson@curtin.edu.au, sampson@unipi.gr

*Corresponding author

ABSTRACT

Project-based learning (PJBL) has been increasingly used to connect abstract knowledge and authentic tasks in educational practice, including computer programming education. Despite its promising effects on improving learning in multiple aspects, PJBL remains a struggle due to its complexity. Completing an authentic programming project involves a complex process of applying programming strategies to design and develop artifacts. Programming strategies are often implicit and hard to capture, but critical for programming performance. This study proposes a visualization-based learning environment that externalizes the complex process of applying programming strategies to the design and development of solutions to an authentic programming project. It aims to make the complex process accessible, trackable, and attainable with the support of technology. Twenty-nine senior college students participated in this study, using the proposed learning environment to complete a PJBL module of ASP.NET. The proposed approach improved students' programming performance and subject knowledge and activated their intrinsic motivation to learn programming.

Keywords

Visualization, Project-based learning, Complex process, Computer programming, Authentic learning, Computer-based learning environment

Introduction

The learning of computer programming has received increased attention with the rapid growth of industry demand for programmers and student interest in programming. However, computer programming is a hard subject for many learners (Govender & Grayson, 2008; Lahtinen, Ala-Mutka, & Järvinen, 2005). A programmer must master both programming knowledge (e.g., concepts, syntax, and semantics) and programming skills and strategies (Robins, Rountree, & Rountree, 2003). The latter refers to the general and specific skills and strategies for integrating abstract programming knowledge into programming tasks by planning, designing, and implementing solutions. Different from programming knowledge, programming skills and strategies are more implicit and harder to capture, yet are critical for programming performance (Soloway, 1986).

Traditional programming education has focused on elaborating abstract knowledge rather than on linking knowledge to specific contexts, which can be labeled as “knowledge driven” or “teacher-centered” education (Robins, Rountree, & Rountree, 2003). As a result, many students demonstrate fragile knowledge (i.e., missing, inert, and misplaced knowledge) and unsatisfactory programming skills and strategies despite passing their programming courses with decent grades. To address the gap, project-based learning (PJBL) has been increasingly promoted by encouraging learners to work with authentic programming tasks and develop artifacts, such as computer programs or design plans, that are realistic products closer to professional reality (Blumenfeld, Soloway, Marx, Krajcik, Guzdial, & Palincsar, 2011).

PJBL contributes to the meaningful learning of abstract knowledge, the improvement of student motivation and confidence, and to the development of communication and problem-solving skills (Kay et al., 2000; Thomas, 2000). Nevertheless, the implementation of PJBL in programming courses, mainly in advanced courses, remains a struggle for many educators (Perrenet, Bouhuisj, & Smits, 2000). PJBL involves a wide range of complex problem-solving activities and extensive hands-on practice, which challenge teachers’ ability to design PJBL curricula and support students throughout their projects. A major concern is that programming tasks involve complex cognitive processes that are inaccessible to learners and instructors. Such complexity can overwhelm learners, making them unable to engage in deeper learning experiences and achieve desired learning outcomes (Helle, Tynijälä, & Olkinuora, 2006; Pucher & Lehner, 2011; Thomas, 2000). Moreover, these complex processes are not always made evident to instructors due to the inherent difficulty of tracking and capturing them, thereby limiting instructors’ capacity to facilitate and improve student performance. Additionally, the resource-consuming nature of PJBL makes it difficult to implement without sufficient time, manpower, and resources (Pucher & Lehner, 2011; Thomas, 2000).
This study attempts to address this challenge by designing a visualization-based learning environment that makes the complex process of carrying out an authentic project accessible, trackable, and attainable throughout the learning process. It uses a web-based learning environment to support PjBL learning with more flexibility in the delivery of learning activities and provision of feedback and support to learners without time and space constraints. This learning environment also has the potential to save costs by reusing learning resources and saving manpower. The study uses ASP.NET as the learning subject, as it is a popular programming course and a mainstream language in the software industry. This study aims to explore the feasibility and effects of the proposed approach in supporting computer programming PjBL. The research questions (RQs) of the study are stated below.

RQ1: How can a visualization-based learning environment be designed to make the complex process of carrying out an authentic project accessible to learners in a programming course?

RQ2: What are the effects of the proposed approach on project-based learning in a programming course?

Literature review

Project-based learning (PjBL)

PjBL is a student-centered pedagogy that encourages students to learn by exploring solutions to real-world problems, mainly by creating artifacts (David, 2008). It highlights the integration of knowing and doing based on the belief that students acquire deep knowledge through active exploration of real-world problems. PjBL is closely related to inquiry learning, problem-based learning, and other learning approaches that aim to fill the gap between learning in a formal instructional environment and the application of knowledge in realistic settings. These learning approaches are underpinned by situated cognition and situated learning theories (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991), which claim that knowledge is rooted in physical and social contexts and that meaningful learning only occurs when knowledge is created and applied.

PjBL distinguishes itself by placing more attention on the development of realistic products closer to professional reality and the assessment of product quality (Adderley et al., 1975; Nuutila, Törmä, & Malmi, 2005). In PjBL contexts, projects tend to involve complex tasks grounded in challenging problems that require students to devise solutions and formulate tangible outcomes by applying their knowledge; learning with a project often takes an extended period of time. Although PjBL has been increasingly promoted in educational practice and has shown promising effects on meaningful learning, it is often not fully implemented (e.g., in programming education) due to its complex and resource-consuming nature (Pucher & Lehner, 2011). Furthermore, the literature has reported inconsistent and inconclusive findings on the effects of PjBL (Blumenfeld et al., 2011; David, 2008).

Learning through complex tasks

Learning through an authentic project is characterized by carrying out a complex task or solving a sophisticated problem, which usually involves complex, implicit processes. The complexity of such processes may generate heavy cognitive loads for learners (Kirschner, Sweller, & Clark, 2006), which is often underestimated by instructors or experts, for whom many of the requisite processes have become largely subconscious or automated. With limited capability to capture the complex processes of authentic tasks, many learners tend to engage in surface instead of deep learning and are unable to achieve the desired learning outcomes (Wang, Kirschner, & Bridges, 2016).

Providing learners with a scaffold or the necessary support to complete a complex task is crucial to learning through ill-structured problems or authentic tasks (Belland, Walker, Kim, & Lefler, 2016; Hmelo-Silver, Duncan, & Chinn, 2007). Scaffolding aligns with the cognitive apprenticeship model, which claims that performing a complex task involves implicit processes. It is critical to make such processes visible for novices to observe, enact, and practice with expert help (Collins, Brown, & Holum, 1991). Approaches to scaffolding for complex tasks have focused on structuring or decomposing a complex task into a set of main actions or key questions to help learners recognize the important goals to pursue in their exploration (Reiser, 2004). The four-component instructional design model presents a guiding framework for systematic learning through complex tasks (van Merriënboer & Kirschner, 2007). The model involves four interrelated components: learning tasks, supportive information, procedural information, and part-task practice. In addition to decomposing a complex learning task, guidance, supportive information, and feedback should be provided for non-recurrent aspects of the tasks; procedural information can be offered as a just-in-time alert for recurrent aspects of the task; and part-task
practice can be used to improve routine skills to reach a required level of automaticity. These suggestions have been incorporated into the design of this study’s proposed learning environment.

Visual representations

Recent research has explored the use of visual representations to externalize and facilitate complex cognitive processes in complex tasks, and has shown improvements in knowledge and task performance (Gijlers & de Jong, 2013; Suthers, Vatrapu, Medina, Joseph, & Dwyer, 2008; Wang, Wu, Kinshuk, Chen, & Spector, 2013; Wu & Wang, 2012). Visual representations or graphic forms (e.g., diagrams, maps, tables, and pictures) have advantages in representing and communicating complex thinking and cognition in flexible ways (Alexander, Bresciani, & Eppler, 2015). By representing information or knowledge both verbally and spatially, such forms can afford more efficient cognitive processing and meaningful communication of complex issues than text messages alone. In recent decades, computer-based visual representations such as concept maps, causal maps, and system models have been increasingly used to extend and augment human cognition, and they have shown their advantages in fostering high-order thinking and meaningful learning in various contexts (Jonassen, 2005).

In computer programming, visual representations and simulation-based visualization tools have been used to visualize the complex structures and algorithms of software programs, explore the development and evolution of programs, and demonstrate the run-time behavior of programs to discover their defects (Koschke, 2003; Roels, Meștereagă, & Signer, 2016). In relation to this, visualization-based learning facilities (e.g., diagrams, pictures, and animations) have been used in programming education to engage learners and assist in their understanding of the abstract concepts and complicated behavior of programs (Eisenberg, Basman, & Hsi, 2014; Hundhausen & Brown, 2007; Naps et al., 2003; Sorva, Karavirta, & Malmi, 2013). They have also been used to help learners integrate separate pieces of knowledge into a coherent whole for meaningful understanding and effective application by allowing them to see a big picture of knowledge on a visual map (Wang, Peng, Cheng, Zhou, & Liu, 2011). Although visualization has shown positive effects on engaging programming learners, research has reported inconclusive findings on its effects on programming learning outcomes, with a major concern for how visualization technology can effectively be integrated with learning and instructional design (Rajala, Laakso, Kaila, & Salakoski, 2008; Sorva et al., 2013).

Methods

This study adopts a design-based research approach, a systematic methodology that creates, builds, and evaluates innovative artifacts or interventions to solve identified problems (Amiel & Reeves, 2008). Design-based research features iterative analysis, design, development, and implementation and close collaboration among researchers and practitioners in real-world settings, leading to contextually sensitive design principles and theories. It is particularly suitable when complex and ambitious educational reform policies are ill specified and the implementation process is uncertain (Wang, Vogel, & Ran, 2011).

Based on relevant learning theories and models, a visualization-based learning environment was designed by decomposing a complex project into a set of main actions, visualizing the process of the project and the process and/or structure of individual actions, and providing specific guidance or strategies for individual actions to make the implicit aspects more accessible. The students used the proposed visualization-based learning environment to complete a PJBL module of ASP.NET. The students’ learning outcomes were measured in terms of subject knowledge, programming performance, and motivation to learn programming. Learner perceptions of the learning environment were also examined, as such perceptions have significant effects on learning in technology-mediated learning environments. Unless a system is properly designed and implemented to the extent that learners find it acceptable, further investigation into the effect of the approach may not produce reliable and meaningful results.

Proposed learning environment

Making the complex implicit process visible

To make the complex and implicit aspects of a programming project accessible, a project is decomposed into a set of main actions based on heuristics or disciplinary strategies. According to the literature on and practice of computer programming (Bassil, 2012; Deek & McHugh, 2002), a programming project typically comprises a
number of key steps, namely problem understanding (or problem formulation), modular design (to design a plan of the solution), process design (to design a detailed solution), coding (to implement a solution), and evaluation and reflection (to test and deliver a solution). Accordingly, the process of completing a programming project is represented in a visual format, as shown in Figure 1. By clicking on the icon of each action, learners can enter the action space for learning and practice. After completing an action, learners can review and refine their outputs.

- Problem understanding
  The first step of a programming project is to formulate a clear understanding of the problem by identifying its goals, givens, and expected results. A structured form is provided to learners to state their understanding of the problem by specifying the requirements and goals of the programming project, as shown in Figure 2.

- Modular design
  A computer program is often organized as a set of functions or modules to be developed independently and then combined to solve the problem. Based on the understanding of the problem, a solution plan can be generated by decomposing the main goals into sub-goals, identifying modular functions to accomplish each sub-goal, and specifying the relationships between the functions. A diagramming tool is provided to learners to build a functional block diagram to outline the plan of the solution, as shown in Figure 3. Moreover, structural analysis and design strategies are offered with a focus on the independence and completeness of the modules.
- **Process design**
  The process within and across the functions must be outlined to illustrate the solution to or algorithm of a given problem, mainly by showing the steps and connections between them. Learners can use the diagramming tool to build a flowchart for the software program, which demonstrates a detailed design of the solution (Figure 3). Learners are also provided with process design strategies, which a focus on priority analysis and critical analysis when designing a complex flowchart involving a number of interactive modules.

- **Coding**
  The modular design and process design can then be translated into program code as a solution to the project. Learners can upload their program codes, which can be reviewed and revised throughout their projects. More importantly, coding strategies are offered to learners with a focus on top-down gradual refinement in addition to their data structures and algorithms.

- **Evaluation and reflection**
  After completing their codes, learners must evaluate their programs by testing and debugging them. Moreover, they can reflect on their performance and areas for possible improvement by reviewing the artifacts generated in each action along with the comments and feedback of their instructor (Figure 4). They can also update their artifacts or solutions and receive further instructor feedback.

![Figure 4. Reflection with feedback](image)

**Providing instructional guidance**

In addition to decomposing a programming project into a set of actions, relevant instructions and guidance on how to apply relevant programming skills and strategies to perform each action are provided to learners as learning materials in the system along with a sample project for illustration. Programming skills and strategies are often implicit and hard to capture, yet are critical for programming performance (Soloway, 1986). Learners need clear guidelines on how to formulate a clear understanding of a problem by identifying its goals and requirements, how to generate a solution plan by organizing a set of modular functions and establishing their relationships, and how to design a logical and appropriate program flow to implement the solution.

**Providing adaptive feedback to individuals**

During the project, the instructor can observe and analyze the artifacts that students generate. Moreover, the instructor can use the system to provide feedback to each student by giving specific comments on his/her problem statement, modular design, program flowchart, and program code. The purpose is to enhance student engagement and in-depth reflection and eventually improve students’ learning outcomes. It is important to make a formative assessment of learning performance based on the artifacts generated not only at the end but also during the project. Visual representations of the learning artifacts play an important role in making previously undetected learning processes accessible for self-reflection by the student and for analysis and feedback by the instructor.

**Participants**

Twenty-nine Year Three college students (12 males and 17 females) participated in this study by using the proposed learning environment to complete a PjBL module of ASP.NET.
Learning task

The participants were asked to complete a representative authentic programming project (membership management) using the proposed learning system. To complete the project, the students went through the main actions of problem understanding, modular design, process design, coding, and evaluation and reflection. The students completed each action by submitting the relevant learning artifacts (i.e., the problem statement, modular design, program flowchart, and code). During the project, the instructor reviewed the students’ learning artifacts and provided individual feedback and comments via the online system.

Procedure

The learning module lasted for 6 weeks. In the first week, the students signed the consent forms for their participation in the study. A questionnaire was then administered to the participants to collect their demographic data. A knowledge test and a programming task were arranged to assess their subject knowledge and programming performance before the study. Next, the participants were given a 30-minute face-to-face instruction session on how to use the proposed system. Relevant information and guidance were also available in the system for flexible access. During the instruction session, a sample project was used for demonstration by the instructor and for self-practice by the students, allowing them to become familiar with the learning environment.

The students started their self-directed learning in the second week. They were asked to complete a project in their free time over a 4-week period. They were also asked to pace themselves and spend 4 hours per week on the project. They could use the online forums for flexible discussion and communication with other students. Based on the log data, most of the students spent approximately 3 hours per week with the system. For each project, most of the students received two to three comments on problem understanding and one to two comments on each of the other parts (i.e., modular design, process design, and coding).

In the sixth week, a knowledge test and a programming task were arranged to assess the students’ subject knowledge and programming performance, respectively. In addition, questionnaires and semi-structured interviews were administered to assess student perceptions and collect feedback.

Measures

Pre-test questionnaire

The pre-test questionnaire was used to collect students’ demographic information and as a self-assessment of their computer skills (very poor, poor, intermediate, good, very good) and intentions to use online learning applications (from strongly disagree to strongly agree).

Post-test questionnaire

The post-test questionnaire was used to collect students’ perceptions of the learning system and intrinsic motivation to learn programming using the proposed system. It used a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Internationally validated and widely adopted measurements from the IT acceptance literature (Arbaugh, 2000; Davis, 1989; Gattiker & Hlavka, 1992) were used to measure student perceptions of the system in terms of usefulness, ease of use, intention to use, and overall satisfaction. Examples of the items include, “The system is useful for my learning,” “The system is easy for me to use,” “I intend to use the system,” and “I am satisfied with the learning system.”

The intrinsic motivation inventory model (Ryan & Deci, 2000) was adapted to measure the students’ intrinsic motivation to learn programming using the proposed system. The motivation measurements involved five scales, including interest/enjoyment, effort/importance, value/usefulness, perceived competence, and pressure/tension. Examples of the items include, “I enjoy attending the learning module,” “I feel confident during the learning module,” and “I get nervous while studying.”
Knowledge tests

Knowledge achievement was assessed before and after the study using two traditional knowledge tests (i.e., a pre-test and a post-test). Different questions of similar difficulty were asked in both tests, which each comprised single-choice, fill-in-the-blanks questions and a short program. All of the test questions were designed by an experienced programming instructor and a programming expert. The total score of each knowledge test was 100.

Programming tasks

Programming performance was assessed before and after the study using two programming tasks (i.e., a pre-task and a post-task). Different tasks of similar difficulty were used in both tests, which were designed by an experienced programming instructor and a programming expert. Both tasks were practical and moderately difficult. For example, in a program, the students were asked to create a class of students, store the name and grades of five courses for each student, calculate the average grade for each student, and display the results of all students.

Based on the literature on programming performance assessment (Deek et al., 1999), the instrument for assessing programming performance consisted of two distinct categories: the programming process and the programming product. The programming process was measured by three subscales: problem understanding, solution planning (i.e., modular design), and solution design (i.e., process design). The programming product (i.e., code) was measured in terms of correctness, efficiency, reliability, and readability. Accordingly, the assessment rubrics for programming performance in this study accounted for four aspects: problem understanding, modular design, process design, and coding. Based on the literature and common practice, the weighting was 40% for coding and 20% for each of the other aspects. The full programming performance score was 20 (8 points for coding and 4 points for each of the other aspects). The detailed assessment rubric with respect to the description, weight, rating criteria, and score range of each aspect is outlined in Table 1.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Weight</th>
<th>Description</th>
<th>Score range</th>
</tr>
</thead>
</table>
| Problem understanding| 20%    | 4 – Problem is clearly and correctly stated. All goals, givens, and results are identified.  
3 – Problem is correctly stated. Most goals, givens, and results are identified.  
2 – Problem is partially stated and/or some facts are identified.  
1 – Problem statement is incorrect and meaningless facts are identified.  
0 – No problem representation/fact identification attempted or completely irrelevant work. | 0 to 4      |
| Modular design       | 20%    | 4 – Detailed and clear planning, with complete goal refinement and task identification.  
3 – Adequate planning, with sufficient goal refinement and task identification.  
2 – Partially correct planning, with some goal refinement and task identification.  
1 – Incorrect planning and meaningless goal refinement.  
0 – No planning/refinement attempted or completely irrelevant work. | 0 to 4      |
| Process design       | 20%    | 4 – Complete module decomposition, organization, and detailed specifications.  
3 – Sufficient module decomposition, organization, and sufficient specifications.  
2 – Partial design and/or some module specifications.  
1 – Improper module decomposition, organization, and specifications.  
0 – No design/specifications attempted or completely irrelevant work. | 0 to 4      |
| Coding               | 10%    | 2 – Correct solution specifications/program code and results consistent with problem requirements.  
1 – Partial solution specifications/program code and/or some results  
0 – No solution specifications/program code, or results inconsistent with problem requirements. | 0 to 2      |
### Efficiency

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Most algorithms, data structures, control structures, and language constructs for this problem situation are appropriate.</td>
</tr>
<tr>
<td>1</td>
<td>Program accomplishes its task, but lacks coherence in choice of either data and/or control structures.</td>
</tr>
<tr>
<td>0</td>
<td>Program solution lacks coherence in choice of both data and control structures.</td>
</tr>
</tbody>
</table>

### Reliability

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Program functions properly under all test cases. Works for and responds to all valid inputs.</td>
</tr>
<tr>
<td>1</td>
<td>Program functions under limited test cases. Only works for valid inputs, but fails to respond to invalid inputs.</td>
</tr>
<tr>
<td>0</td>
<td>Program fails under most test cases.</td>
</tr>
</tbody>
</table>

### Readability

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Program includes commented code, meaningful identifiers, indentation to clarify logical structure, and user instructions.</td>
</tr>
<tr>
<td>1</td>
<td>Program lacks clear documentation and/or user instructions.</td>
</tr>
<tr>
<td>0</td>
<td>Program is totally incoherent.</td>
</tr>
</tbody>
</table>

---

**Semi-structured interview**

The interviews collected the students’ comments and feedback on the learning module by requiring students to write the responses on the paper. Each interview addressed two open-ended topics: (1) the strengths and weaknesses of the learning module and (2) suggestions for improving the learning module.

### Results

Twenty-nine students completed the study. Most of them had an intermediate level of computer skills and a neutral intention to use online learning applications.

### Questionnaire

#### Perceptions of the learning environment

The participants reported positive perceptions of the learning system in terms of its usefulness (Mean = 4.12, SD = .46), their intentions of using it (Mean = 4.22, SD = .54), and their overall satisfaction (Mean = 4.14, SD = .57). However, their perceptions of the ease of use of the system were weakly positive (Mean = 3.66, SD = .58). An internal consistency analysis using Cronbach’s alpha confirmed that all of the subscales were reliable (.75 for usefulness, .74 for ease of use, .70 for intention to use, and .74 for overall satisfaction).

### Motivation to learn programming

The participants reported having a clear motivation to learn programming using the proposed learning environment. The means were 4.09 (SD = .46) for interest/enjoyment, 4.39 (SD = .44) for value/usefulness, 3.89 (SD = .57) for effort/importance, 3.63 (SD = .55) for perceived competence, and 2.95 (SD = .86) for pressure/tension. As shown by the data, the students had a strong interest in learning using the proposed system and found the learning module highly useful. Moreover, they reported that they had made efforts in the learning task and felt competent in completing the project. Their perceived pressure during the learning process was slightly lower than neutral. An internal consistency analysis using Cronbach’s alpha confirmed that all of the subscales were reliable (.76 for interest/enjoyment, .77 for effort/importance, .75 for value/usefulness, .79 for perceived competence, and .86 for pressure/tension).
Knowledge tests

The descriptive statistics for the knowledge-test scores and the paired sample t-test results for the difference between the pre-test and post-test are presented in Table 2. The students demonstrated significantly improved programming knowledge after completing the learning module.

**Table 2.** Descriptive statistics and paired sample t-test for knowledge achievement (n = 29)

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Paired sample t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Mean</td>
<td>46.34</td>
<td>17.29</td>
<td>53.31</td>
</tr>
</tbody>
</table>
| Note. **p < .001.**

Programming tasks

Two domain experts assessed the students’ performance on the programming tasks, and their scores were averaged. The two raters were blind to student identification and test information (i.e., whether the test was pre-test or post-test). The inter-rater reliability was computed using Cohen’s kappa. The results exceeding .8 indicated perfect agreement, while .6 indicated substantial agreement. The results were .872 for problem understanding, .866 for modular design, .815 for process design, and .675 for coding (all significant at the .001 level), reflecting perfect agreement and consistency between the raters on the first three scales and substantial agreement on the last scale.

The descriptive statistics for the students’ programming performance before and after the study and Wilcoxon signed-rank test for the difference between the two are presented in Table 3. The students showed significant progress on all of the programming performance scales.

**Table 3.** Descriptive statistics and Wilcoxon signed-rank test for programming performance (n = 29)

<table>
<thead>
<tr>
<th>Programming performance</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Paired sample t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Problem understanding</td>
<td>2.71</td>
<td>.921</td>
<td>3.03</td>
</tr>
<tr>
<td>Modular design</td>
<td>2.57</td>
<td>.904</td>
<td>3.345</td>
</tr>
<tr>
<td>Process design</td>
<td>2.41</td>
<td>.814</td>
<td>3.293</td>
</tr>
<tr>
<td>Coding</td>
<td>5.67</td>
<td>1.611</td>
<td>6.29</td>
</tr>
<tr>
<td>Total score</td>
<td>13.36</td>
<td>3.076</td>
<td>15.97</td>
</tr>
</tbody>
</table>
| Note. */p < .05; **p < .01; ***p < .001.**

Semi-structured interview

The comments shared by the students in their responses are presented in Table 4. During the interview, many students reported that the proposed learning module was highly useful in improving their programming thinking and problem-solving skills and in supporting their self-directed learning. Some of the students mentioned that they enjoyed the practical and meaningful learning experience and felt strongly motivated and interested in learning computer programming. Meanwhile, the students reported experiencing difficulties completing the learning tasks and suggested having more interactions with other students during the learning process. This is achievable not only through online forums but also through group-based learning with face-to-face communication. Some of the students suggested adding more learning resources and programming tasks to the system for self-study. The students also suggested improving some interfaces or functions of the system to make the system more attractive and easy to use.

**Table 4.** Students’ responses to semi-structured interview questions

<table>
<thead>
<tr>
<th>Students’ comments</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strengths of the learning module</strong></td>
<td></td>
</tr>
<tr>
<td>Improving my programming thinking</td>
<td>21</td>
</tr>
<tr>
<td>Supporting self-directed learning or improving my self-directed learning capability</td>
<td>21</td>
</tr>
<tr>
<td>Enhancing my programming problem-solving skills</td>
<td>16</td>
</tr>
<tr>
<td>Offering practical and meaningful learning experiences</td>
<td>8</td>
</tr>
<tr>
<td>Insightful and useful support afforded by the learning system</td>
<td>8</td>
</tr>
<tr>
<td>Activating my interest and motivation for learning programming</td>
<td>6</td>
</tr>
</tbody>
</table>
Weaknesses of the learning module

- Insufficient interactions with peers: 16
- Inadequate learning resources: 9
- Learning tasks are too difficult: 8
- Not comfortable with self-directed learning: 7
- Some interfaces or functions of the learning system are not attractive or easy to use: 6

Suggestions for improvement

- Allowing more interactions between students, probably via group-based learning: 20
- Including more learning materials and programming tasks in the system: 18
- Adjusting the difficulty level of the learning projects: 10
- Making some interfaces or functions of the learning system more user-friendly: 5

Discussion

RQ1: How can a visualization-based learning environment be designed to make the complex process of carrying out an authentic project accessible to learners in a programming course?

The proposed learning environment was designed by decomposing a complex project into a set of main actions, visualizing the process of the project and the process and/or structure of individual actions, and providing relevant guidance or strategies for individual actions to make the implicit aspects more accessible. This made the complex process of PjBL visible and accessible to learners and instructors, empowered the students to engage in deep learning and reflection, and enabled the instructors to observe individual performance and provide adaptive feedback and support throughout the project.

RQ2: What are the effects of the proposed approach on project-based learning in a programming course?

The results show the promising effects of the proposed learning environment as reflected by students’ achievements in subject knowledge and programming performance and by their perceptions and motivations to learn.

First, the participants made significant improvements on programming task performance and subject knowledge after completing the learning module. Their progress in programming performance was significant in all aspects (i.e., problem understanding, modular design, process design, and coding).

Second, the students reported positive perceptions of the learning system in terms of its usefulness, their intentions of using it, and their overall satisfaction. However, their perceptions regarding the ease of use of the system were weakly positive. The participants also reported having a clear motivation to learn programming using the proposed learning environment. In particular, they enjoyed the learning experience and found it highly useful. They made efforts on the learning task and felt competent in completing the project, although they perceived some pressure during the learning process.

Third, the interview results confirm the students’ perceptions and motivations to learn in addition to their comments on the strengths of the learning environment in facilitating self-directed learning. The students also provided suggestions for improving the learning module by enabling more peer interaction during the learning process, providing more online learning resources, and making the system more user-friendly. These issues will be addressed in a future study following the design-based research paradigm.

The findings align with those of related studies on PjBL (Blumenfeld et al., 2011) and visual representations in programming education (Sorva et al., 2013; Wang, Peng, Cheng, Zhou, & Liu, 2011). It should be noted that the visual representations used in previous studies have mainly targeted the learning of programming knowledge (e.g., basic concepts, syntax, and semantics) rather than programming skills and strategies, which are harder to capture yet are more critical for programming performance. Although specialized visual representations, such as functional block diagrams and flowcharts, are more applicable to learning programming skills and strategies (Blackwell, Whitley, Good, & Petre, 2001), it is not easy for students to use such tools to generate a satisfactory solution to a program. Research is needed to explore how visualization technology can effectively be integrated with programming learning and instruction (Rajala et al., 2008; Sorva et al., 2013). The findings of this study contribute to the efforts to address the concern by visualizing the complex process of performing an authentic programming project using relevant strategies.
Conclusion

PiBL is a promising approach to learning that connects abstract knowledge to authentic tasks. It has been increasingly promoted in educational practices, including computer programming education. Although written exams and program codes have been widely used for summative assessments in programming courses, grades may not reflect students’ actual programming performance, as many students with good grades still have many problems completing actual programming tasks. PiBL has the potential to address this gap. Nevertheless, making the complex process of authentic projects accessible for effective learning and instruction is a pressing issue that makes it difficult to fully implement PiBL and to achieve the desired learning outcomes. Although research has shown the promising effects of visualization technology in programming education, the literature has reported inconsistent findings concerning how visualization technology can effectively be integrated with learning and instruction (Rajala et al., 2008; Sorva et al., 2013).

In an attempt to address this challenge, this study proposes the design of a visualization-based learning environment that externalizes the complex process of applying programming strategies to the design and development of solutions to an authentic programming project. It aims to make the complex process accessible, trackable, and attainable with the support of technology. In particular, it engages students in deep learning and reflection on the process and enable instructors to observe individual performance and provide adaptive feedback throughout the project. A learning environment focused on decomposing a complex project into a set of main actions, visualizing the project process and the process and/or structure of individual actions, and including relevant guidance or strategies for individual actions was designed. After using the proposed system to complete a PiBL module of computer programming, the students showed significant pre-post improvements in subject knowledge and programming performance. They also reported positive perceptions of the learning environment and favorable motivation and confidence to learn programming using the proposed approach. Following the design-based research paradigm, the proposed learning environment will be refined based on the results of the study, especially the comments and feedback of the participants. A control group design will be carefully implemented in a future study for further analysis.

The findings contribute to the knowledge of how complex PiBL can be adequately implemented through the effective design of technology-enhanced environments, particularly by making the complex aspects of PiBL visible and accessible for deep learning and reflection and by making the learning artifacts generated in the process available for timely analysis and feedback. Capturing and facilitating the complex process of carrying out an authentic project is a challenging issue in PiBL. The proposed visualization approach has the potential to gain holistic understandings of learner activity, cognitive processes, and performance in learning with authentic projects or complex tasks to improve learning and performance in such contexts. Although this study pertains to the domain of computer programming, the proposed approach has the potential to be transferable to PiBL in other disciplines.

Acknowledgements

This research is supported by the General Research Fund (Project No. 17201415) from the Hong Kong SAR Government, and the Seeding Fund for Basic Research (Project No. 201611159071) and the Faculty Research Travel Support for Full-time MPhil and PhD students from the University of Hong Kong. It was partially supported by the Greek General Secretariat for Research and Technology under the Matching Funds 2014-2016 (Project No. 325123). The authors would thank Professor Haijing Jiang for his valuable support for this project.

References


286


Can Students Identify the Relevant Information to Solve a Problem?

Lishan Zhang1,2, Shengquan Yu1,2*, Baoping Li1,2 and Jing Wang1

1Beijing Advanced Innovation Center for Future Education, Beijing Normal University, China // 2Faculty of Education, Beijing Normal University, China // lishan@bnu.edu.cn // yusq@bnu.edu.cn // libp@bnu.edu.cn // jingwang0422@163.com

Corresponding author

ABSTRACT

Solving non-routine problems is one of the most important skills for the 21st century. Traditional paper–pencil tests cannot assess this type of skill well because of their lack of interactivity and inability to capture procedural data. Tools such as MicroDYN and MicroFIN have proved to be trustworthy in assessing complex problem-solving performances in the dynamic environments representing linear structural equations and finite state automata. In contrast to previous studies, this paper introduces a system that assesses how an individual acquires information to solve real-life problems. Specifically, the system investigated whether fifth-grade students could recognize a situation in which additional information is needed, acquire the relevant information, and finally apply it to solve their problems. By running an experiment with a total of 32 fifth-grade students, we found that the students were usually able to recognize situations when they needed additional information. However, students sometimes spent too much time reading irrelevant materials, which was significantly correlated with worse problem-solving performance \[ r = 0.417, p = .018 \].

Keywords

Problem solving, Problem-based learning, Information identification

Introduction

Regardless of their occupation, people need to handle different types of problems every day. Real-life problems are usually very complex and cannot be solved in a routine manner. Therefore, knowing how to solve these unusual problems has become an essential skill for the 21st century (Greiff et al., 2014a; Griffin, McGaw, & Care, 2012; Neubert, Mainert, Kretzschmar, & Greiff, 2015). Problem solving is not only a skill to deal with real-life situations, but also plays an important role in many learning environments, such as problem-based learning (PBL) (Merritt, Lee, Rillero, & Kinach, 2017).

Problem solving is the process of finding a method to achieve a goal from an initial state. Depending on the domain, the initial state, goal, and means can be very different. Therefore, domain expertise usually plays a dominant role in an individual’s problem-solving performance, as described by Chi and Glaser (1983). Well-educated adults may exhibit equally good domain-general problem-solving strategies. However, young students clearly have different levels of competence in solving complex domain-general problems (Findings, 2014), which has fostered a great demand for teaching domain-general problem-solving skills to these students (Greiff et al., 2014a).

In addition, a domain-general problem-solving ability is essential for PBL, which is an effective method for enhancing deep learning (Dolmans, Loyens, Marcq, & Gijbels, 2015). In PBL, students are expected to learn by solving open-ended problems. Open-ended questions are usually complex and poorly structured; therefore, students should have strong self-regulation abilities to be successful in this type of learning. In other words, it is not appropriate to assume that every student is well prepared for PBL. Thus, PBL often involves extensive tutor facilitation, which some educators find difficult and frustrating (Wood, 2003). When facilitation is absent or insufficient, PBL is sometimes found to be less effective than traditional lecture learning (Kirschner, Sweller, & Clark, 2006). Recently, many big cities in China, such as Beijing and Shanghai, have developed a strong trend of adopting PBL in elementary schools. However, as the normal size of a Chinese elementary school class is 40 students, it is not possible for a teacher to help all students efficiently. Therefore, it is even more important to assess students’ abilities in conducting PBL-related activities in China to enable teachers to have a better sense of which of their students may need the most help.

Our main objective in problem-solving assessment is to check whether students are ready for PBL; thus, we need an assessment tool fit for this purpose. Some tools have been developed for domain-general problem-solving assessment. Among these tools, MicroDYN and MicroFIN (Schweizer, Wüstenberg, & Greiff, 2013) are the best established. When using these tools, students are required to investigate the complex dependencies of several variables within a dynamically changing situation. Indeed, these tools can be used to reliably assess problem-solving abilities by describing problem situations with linear structural equations and finite state automata.
However, PBL also requires other dimensions of problem-solving ability. Previous studies have shown that students often failed in PBL because of its high cognitive load (Kirschner, Sweller, & Clark, 2006; Sweller, 1988; Tuovinen & Sweller, 1999). One important of cognitive load-related factor is reading literacy. PBL usually involves a great deal of information searching and selection, which requires students to be able to decide what they need to know to resolve their problem (Holliday, 2006). In this context, students need to do much more discontinuous reading than continuous reading. However, Chinese students exhibit worse performance in discontinuous reading than students from other countries according to the report from the 2009 Programme for International Student Assessment (PISA) (Organisation for Economic Co-operation and Development (OECD), 2010). We are concerned that many Chinese elementary school students may experience difficulties in actively searching for the relevant information to solve their problems. Thus, we built an assessment system to evaluate our students’ problem-solving ability and their ability to pay attention to different types of information. We will introduce this system in this paper and report how students’ attention to information is related to their problem-solving performance. By conducting this experiment, we mainly wanted to answer two research questions:

RQ1. Are students able to source the relevant information and apply them in solving real-life problems?

RQ2. Can the attention given to some specific types of information be used as a predictor of problem-solving performance?

The paper is organized as follows. We review the related studies and then introduce our system and the assessment task. We then describe the experimental design and report the results. Finally, we discuss our findings and conclude with some final remarks.

Related work

Complex problem solving in dynamic environments

While being assessed in a simulated dynamic environment, individuals’ core competence for problem solving is shown in their ability to determine the complex dependencies among the observable variables. The most well-known project in this field is probably MicroDYN, which was developed by Schweizer et al. (2013), where students are expected to determine the dependencies of variables in a complex system by manipulating the variables and observing their effects in a dynamic environment. A recent study found that the assessed skill has a strong correlation with traditional reasoning test performances, but was also an independent dimension of ability (Kretzschmar, Neubert, Wüstenberg, & Greiff, 2016; Greiff & Neubert, 2014b). Previous studies have shown the value of this perspective. However, this perspective is not sufficient to explain all types of problem-solving activities. Many problems do not contain complex variable dependencies, but instead need students to distinguish related information from a great number of documents. To solve these problems, problem solvers must be clear about what information they want and purposively search for this information.

Guidance required for problem-based learning

Both problem- and project-based learning can be abbreviated as PBL. However, PBL is used to abbreviate problem-based learning in this paper. PBL has attracted research attention for many years (Holliday, 2006; Kirschner, Sweller, & Clark, 2006; Merritt, Lee, Rillero, & Kinach, 2017) and has been praised because it can motivate students and trigger deep thinking (Merritt, Lee, Rillero, & Kinach, 2017). However, PBL has simultaneously received much criticism (Holliday, 2006; Klahr & Nigam, 2004; Kirschner, Sweller, & Clark, 2006; Patel, Groen, & Norman, 1993; Tuovinen & Sweller, 1999). The main issue for PBL is that students need assistance and facilitation during the learning process and have difficulty in self-regulating their learning. Students, especially those with low levels of prior knowledge, can easily develop a great burden on their cognitive load (Sweller, 1988) and fail to distinguish what they really need to know and learn. This may result in students learning useless or even incorrect knowledge and concepts (Harris & Graham, 1994). Therefore, teachers need to be highly involved in PBL to motivate, regulate, and provide hints to their students so that they do not get lost in looking for the relevant information. Recently, researchers have started to integrate PBL with direct instruction, which seems to be a viable solution (Holliday, 2006; Jalani & Lai, 2015). Direct instruction usually teaches learning strategies for students at the metacognitive level and helps them to avoid making errors in their studies. The direct instruction content is mainly made based on the teacher’s experience. By developing this assessment tool, we hope to provide teachers with a better sense of what should be taught to their students.
Typical problem-based learning practice

PBL is most widely used in medical education; therefore, this domain is used as an example to review how PBL may be implemented in practice. Students are usually given a patient’s problem related to the skills to be taught. They then learn the related material through self-directed studies to solve the patient’s problem (Distlehorst, Dawson, Robbs, & Barrows, 2005). This learning process often involves group study and discussion. The group size can be varied by cases, but usually does not exceed eight students. Students are sometimes grouped by their interests so that they can search for the problem-related information based on their preferences (Distlehorst, Dawson, Robbs, & Barrows, 2005). Teacher facilitation is provided to help students find the relevant information and improve group discussions (McParland, Noble, & Livingston, 2004). In terms of assessment, the studies usually focus on information acquisition, self-regulation, and collaborative study (Distlehorst, Dawson, Robbs, & Barrows, 2005). Teachers usually conduct the assessment by grading students’ submitted reports. Dickson et al. (2016) assessed nursing clinical judgement via a computer-simulated environment by analyzing the students’ recorded behaviors. In that system, students can gather information by looking at their patient’s temperature, medical report, lab results, and vital signs. The students then need to determine the appropriate treatment based on the gathered information. Therefore, the identification of relevant information is one of the most important steps in PBL.

Behavioral analysis

An evidence-centered design (ECD) should be adopted to create a system to support the analysis of students’ intentions based on behavioral data. The fundamental design concept was first described by Mislevy (1994). An ECD defines an assessment framework to ensure that evidence is gathered appropriately to be able to interpret the underlying purpose of the assessment. Many common design features are shared by tutoring and assessment systems, although they may be implemented based on their own interpretation of an ECD (Shute, Wang, Greiff, Zhao, & Moore, 2016; Shute, 2011). The very first aspect of adopting this framework is to define the domain modeling, i.e., to clarify the skills to be assessed and sketch the relationships among the proficiencies of the skills, tasks, and evidence. The next main step is to detail the relationships. This process can be factored into three models: student, evidence, and task models. When ECD is applied in an interactive environment-based assessment, a task model defines the story line and how the tasks can elicit students to interact (Halverson & Owen, 2014). An evidence model then describes how the interactions should be analyzed. The evidence model can be based on either statistics or rules. For example, Zhang et al. (2014) inferred students’ proficiencies in a meta-strategy using a set of rules and a special sequence of behaviors. Using a hidden Markov model, Schwartz et al. (2009) explored how students interacted with teachable agents. Bayesian networks are widely known statistical analysis models used in many tutoring systems (Almond, Mislevy, Steinberg, Yan, & Williamson, 2015). Our system adopts an ECD framework that mainly uses rules to analyze students’ problem-solving competences.

The assessment system

To succeed in PBL, we expect that given a real-life situation, students can at first identify what factual knowledge they need to know to solve the corresponding problem, and then search for and solve the problem by using the relevant factual knowledge. We provided problem-irrelevant and -relevant information to the students to evaluate how well they could deal with the related cognitive load. The relevant and irrelevant information were presented together, but different types of information were put into separate documents. Skilled students are expected to keep their cognitive load low by focusing only on the task-related materials. Because the students’ patterns of accessing different types of information can reflect how they value the information (Johnson, Häubl, & Keinan, 2007), we inferred the students’ focus by observing their recorded behavioral patterns. Therefore, we designed and implemented the system in such a way that the students’ information-seeking and problem-solving behaviors can be recorded and detected easily. The system includes one assessment task containing four test items. The relevant and irrelevant documents are stored in a virtual library, which is named the “materials center.” Students must visit the materials center and read the relevant materials to solve the related problems. We are interested in how students value different information in the materials center. The most straightforward method is to directly ask students to select the information that they think is important. However, this design will interrupt the students’ natural problem-solving processes. Thus, we non-intrusively infer how they value the information by analyzing the log files. The rest of this section discusses the user interface, details the contents of the assessment task, and describes the types of behaviors that we recorded.
The user interface

To support the structure of the test items, the system provides some common functionalities and utilities. To allow students to access the relevant materials conveniently, a navigation bar is placed at the right-hand side of the area where the test items are displayed (Figure 1). All associated materials for the current assessment task are stored in the component, “materials center,” which is located in the navigation bar. The materials center contains not only the relevant materials but also some irrelevant ones. Therefore, the follow-up data analysis could reveal how students chose different materials. As soon as a student clicks on the materials center, a new window pops up that shows a list of the materials (Figure 2, left). Each entry in the list shows both the title and a brief description of the document. When a student clicks on the name of the document, its details are displayed (Figure 2, right). The student can click on the back button to return to the list panel and access other materials. Students can also close the materials center whenever they want by clicking on the close button. High-achieving students are supposed to be very clear about what they are looking for and can thus locate the relevant document quickly. By comparison, low-achieving students could flounder and may randomly access many irrelevant documents. As soon as the students finish a test item, they can proceed to the next one.

The assessment task

To help the students learn the user interface, they must perform an introductory task before performing the actual assessment task. The introductory task is very simple: students need to calculate how much time would be taken to fly from Beijing to Washington, DC. The airplane speed has been given in the description of the problem. The distance between the two cities is included as a document in the materials center, which also includes other irrelevant material. The introductory task asked students to first locate the relevant material, and then perform...
the calculation. The students must finish both parts of the introductory task to complete it. Otherwise, the system prompts the students to recognize which part they did not finish.

<table>
<thead>
<tr>
<th>Table 1. Summary of test items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item name/Feature</strong></td>
</tr>
<tr>
<td>Ticket purchase</td>
</tr>
<tr>
<td>Tent capacity calculation</td>
</tr>
<tr>
<td>Tent assignment</td>
</tr>
<tr>
<td>Food and water supply</td>
</tr>
</tbody>
</table>

The actual assessment task contains four test items and uses camping as the story line. The four standalone test items share the same background story line. The features of these four test items are summarized in Table 1. In addition to the traditional types of items, such as multiple choice and fill-in-blank, students need to interact with a simulated environment to solve problems in two test items. These items are “interactive.” For each test item that requires students to search for relevant information, we provided three types of materials at three different levels of relevance: strong, medium, and weak. Students are required to obtain materials with strong relevance to solve the problem. Materials with medium relevance are related to the problem situation and materials with weak relevance are completely irrelevant to the problem. Some material examples are listed in Table 2. The remainder of this section describes each test item in detail.

<table>
<thead>
<tr>
<th>Table 2. Test items and their materials with strong, medium, or weak relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test item</strong></td>
</tr>
<tr>
<td>Ticket purchase</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tent capacity</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Food and water supply</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Test item 1: “Ticket purchase”**

In the first test item, students are expected to read the relevant materials to determine the best way to purchase tickets. Students are told that they plan to go camping outside. They must fly to the camping area because it is far from where they are living. Students are required to decide whether they should purchase group or individual tickets. The problem clearly describes the flight destination. Students should recognize that they need to distinguish the difference between the two purchasing methods, visit the materials center to check the ticket-purchasing rules, and then finally make their decision.

**Test item 2: “Calculating tent capacity”**

The second test item requires students to calculate the capacity of their tents (Figure 3, left). The students are expected to measure the sides of three tents correctly and then use the methods described in the relevant materials to calculate the capacities of the tents, which differ only in their sizes. To measure a side of a tent, the students need to click on the side, read the length from a pop-up image (Figure 3, right), fill out the value, and close the pop-up image. Each tent has five sides to be measured, but only two of them are useful for the calculation of the capacity. Therefore, the students can potentially make some invalid interactions—i.e., making unnecessary measurements—in this task. Students who read and understood the relevant materials well should be able to avoid performing these invalid interactions. Although making invalid interactions does not harm the students’ final performance score, it is a sign of shallow thinking.
Figure 3. Screenshot of test item 2: “Tent capacity calculation” (left), pop-up image showing the tent side measurement (right)

Test item 3: “Tent assignment”

The third test item does not require students to read any relevant materials. However, the students can visit the materials center if they want. To solve this problem, students need to assign 54 persons to 7 tents (Figure 4). The 54 persons can be categorized into four types, according to their age and gender: i.e., boys, girls, and male and female adults. There are three different types of tents. Each type of tent can contain a limited number of persons. The tent assignment should satisfy three constraints: (1) People in the same tent must be the same gender; (2) There must be at least one adult in a tent; (3) The number of persons in a tent cannot exceed its limit. This problem checks whether students can recognize a situation where they do not need to refer to additional material.

Figure 4. Screenshot of test item 3: “Tent assignment”

Test item 4: “Food and water supply”

In the last task, students are told that they can only resupply their food and water at the end of each day during their camping trip; therefore, they need to calculate how much food and water they should prepare for each day. The materials center provides the corresponding documents that describe how much food and water an adult usually needs to consume daily. The students must obtain this information to solve the problem. In contrast to the first test item, the students need to type their answer. Because the answers are specific numbers, they can be easily graded automatically.

Behavioral record

Log files have been widely used to understand students’ performance (Greiff, Wüstenberg, & Avvisati, 2015; Kuo & Wu, 2013). Our system records every single interaction, such as clicking on an alternative and accessing the relevant materials. Thus, the behavioral data can be analyzed offline. When a test item is a multiple-choice question, the set of possible interactions is limited and only contains “choosing an alternative” and “accessing different types of materials.” However, the number of possible operations a student can make in an interactive test item is potentially much greater and completely depends on the specific problem. To standardize all interactions, except for materials center-accessing behaviors, they are labeled with four different types: correct, incorrect, valid, and invalid. Correct and incorrect behaviors describe the correctness of an interaction, but the
correctness criteria are changed by specific problems. Test item 2 intentionally embedded some unnecessary interactions towards the problem goal. Students can either skip or finish these steps. In test item 2, a necessary interaction is labeled as “valid” and other interactions are labeled as “invalid.” For each material-accessing behavior, we recorded the timestamp and the material’s corresponding level of relevance. The details of our analysis are described in the next section.

**Experiment design and analysis model**

**Experiment design**

All students who participated in our experiment were in the fifth grade. We performed a pilot study with approximately 20 students in the same grade to evaluate the usability of the system and the assessment task. The introductory task was not included initially. As a result, the students in the pilot study had no idea how to use the materials center to help them solve problems. Thus, we created an introductory task to guide the students in accessing the materials center. This introductory task essentially helped students to become familiar with the user interface. The pilot study also helped us identify several software bugs and ensure that the problem description could easily be understood by fifth-grade students. Students were required to finish the introductory task before starting the actual assessment task. They were asked to finish the assessment individually in a class.

**Analysis model**

The first question we needed to answer was how well students performed when they had to find out part of the facts related to the problems by themselves. Therefore, we calculated the descriptive results of their performance scores on each individual test item. In addition, we wanted to understand how students solve problems by observing how they used the materials center and identify their issues during the problem-solving process. Therefore, we mined the relationship between students’ behavioral patterns and performance outcomes on test items. Pearson’s correlation coefficient was calculated to measure the relationship. Linear regression was further applied to evaluate whether the students’ behavioral patterns could be used to predict their final problem-solving performance.

In analyzing the students’ behavioral patterns, we focused on their interactions with the documents in the materials center. Because we categorized all documents into three types for each test item—strong, medium, and weak—we could examine how students allocate their attention to the three types of materials. As soon as students opened a document, the pop-up window would prevent them from performing all the other behaviors (e.g., answering the question or opening another document). We used the time that the student stayed at an opened document to reflect how important the student thought the material was. We assumed that if the students spent at least 5 seconds in reading a document, they probably considered the material as somewhat important. Each reading behavior was labeled as either short-reading or non-short-reading behaviors. Five seconds was used as the threshold to distinguish between these behaviors. Therefore, we classified the reading behaviors into 2 (short-reading and non-short-reading behaviors) × 3 (strong, medium, and weak relevance) = 6 different types.

In addition to the test item scores, the test item “Tent capacity calculation” could also report other information to reflect the students’ problem-solving performance. As mentioned earlier, the students can potentially perform invalid work in solving their problems. Specifically, an example of invalid work is measuring tent sides unnecessarily. Because the related information in the materials center clearly states which two sides should be measured to calculate the capacity of tents, the students who acquire the information accurately and are thinking deeply should be able to avoid invalid work (Holliday, 2006). Thus, the percentage of unnecessary sides measured was calculated to reflect whether the students were thinking shallowly.

**Results**

**Descriptive results**

As mentioned above, there were four test items in total. Each item had a weight of 1 point. Therefore, the maximum possible score on the test was 4. Thirty-two fifth-grade students participated in the experiment. The average score on the entire assessment task was 1.572 ($SD = 0.762$). Table 3 reports students’ performance and their reading behaviors on each individual test item. Because the test item, “Tent assignment,” does not have any
relevant materials, we aggregated the students’ performance and their reading behaviors on the other three test items (Table 3). The results indicated that students had low scores while solving the problems that required reading relevant materials, and scored relatively high while solving the problems without relevant materials.

**Table 3. Descriptive results of the test items**

<table>
<thead>
<tr>
<th>Task</th>
<th>Ticket purchase</th>
<th>Tent capacity calculation</th>
<th>Tent assignment</th>
<th>Food and water supply</th>
<th>Overall performance without test item 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>0.5</td>
<td>0.25</td>
<td>0.68</td>
<td>0.136</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>(SD = 0.508)</td>
<td>(SD = 0.237)</td>
<td>(SD = 0.413)</td>
<td>(SD = 0.195)</td>
<td>(SD = 0.572)</td>
</tr>
<tr>
<td>Number of short-reading</td>
<td>0.063</td>
<td>0.031</td>
<td>N/A</td>
<td>0.094</td>
<td>0.19</td>
</tr>
<tr>
<td>relevant materials</td>
<td>(SD = 0.242)</td>
<td>(SD = 0.174)</td>
<td></td>
<td>(SD = 0.291)</td>
<td>(SD = 0.464)</td>
</tr>
<tr>
<td>Number of short-reading</td>
<td>0.094</td>
<td>0.16</td>
<td>N/A</td>
<td>0.063</td>
<td>0.31</td>
</tr>
<tr>
<td>medium relevant materials</td>
<td>(SD = 0.291)</td>
<td>(SD = 0.44)</td>
<td></td>
<td>(SD = 0.242)</td>
<td>(SD = 0.527)</td>
</tr>
<tr>
<td>Number of short-reading</td>
<td>0.125</td>
<td>0.094</td>
<td>0</td>
<td>0.59</td>
<td>0.81</td>
</tr>
<tr>
<td>weak relevant materials</td>
<td>(SD = 0.415)</td>
<td>(SD = 0.291)</td>
<td>(SD = 0)</td>
<td>(SD = 2.16)</td>
<td>(SD = 2.17)</td>
</tr>
<tr>
<td>Number of non-short-</td>
<td>0.44</td>
<td>0.75</td>
<td>N/A</td>
<td>0.50</td>
<td>1.69</td>
</tr>
<tr>
<td>reading strong relevant</td>
<td>(SD = 0.556)</td>
<td>(SD = 0.612)</td>
<td></td>
<td>(SD = 0.66)</td>
<td>(SD = 0.982)</td>
</tr>
<tr>
<td>materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of non-short-</td>
<td>0.56</td>
<td>0.28</td>
<td>N/A</td>
<td>0.094</td>
<td>0.94</td>
</tr>
<tr>
<td>short-reading medium</td>
<td>(SD = 0.788)</td>
<td>(SD = 0.514)</td>
<td></td>
<td>(SD = 0.291)</td>
<td>(SD = 0.899)</td>
</tr>
<tr>
<td>relevant materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of non-short-</td>
<td>0.125</td>
<td>0.094</td>
<td>0.15</td>
<td>0.34</td>
<td>0.56</td>
</tr>
<tr>
<td>short-reading weak</td>
<td>(SD = 0.545)</td>
<td>(SD = 0.291)</td>
<td>(SD = 0.330)</td>
<td>(SD = 0.988)</td>
<td>(SD = 1.09)</td>
</tr>
<tr>
<td>relevant materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Correlation results**

Pearson’s correlation coefficients were calculated to determine whether material-reading behaviors were related to student performance. Specifically, for each student, we first aggregated his/her total number of short-reading and non-short reading behaviors on materials with strong, medium, or weak relevance. We then calculated the correlation between the six factors and the students’ overall performance except for the task, “Tent assignment.” The results are shown in Table 4. The amount of non-short-reading behaviors for materials with weak relevance was the only factor that significantly correlated with problem-solving performance. All the factors were also used to construct linear regression to predict overall performance without test item 3 using a stepwise algorithm. We found that the overall performance can be predicted by the number of non-short-reading weak relevant materials ($\beta = -0.417$, $R^2 = 0.174$) and a constant.

The percentage of unnecessary sides measured in the task, “Tent capacity,” was used to reflect students’ shallow thinking. The correlation between material-reading behaviors and evidence of shallow thinking was calculated. Among all of the reading behaviors, the number of short-reading behaviors for materials with weak relevance was the only one that significantly correlated ($r = 0.398$, $p = .024$) with evidence of shallow thinking. This means that students who quickly scanned a lot of irrelevant materials also tend to simply do work without thinking about the reason.
The results showed that the students did not perform very well in general. Although this type of assessment task is very different from those that they are usually assigned, these problems themselves are not very difficult. Teachers believe that fifth-grade students have enough prior knowledge to solve them. Probably because the problems were somewhat ill-structured—i.e., part of the related factual knowledge was stored in the materials center—and students sometimes failed to locate the information precisely. Most students could not answer the questions correctly. This claim is supported by our data analysis. The more students spent their time in reading completely irrelevant materials, the more likely these students would receive a bad score on their overall performance. This result is consistent with the theory of cognitive load (Sweller, 1988). Students should efficiently use their limited working memory to handle challenging problems. Reading and processing irrelevant information might occupy too much of students’ working memory during problem solving, and increase their cognitive loads. As PBL will make students explore an even bigger problem space than the one in the experiment, teachers should provide strong facilitation to guide their students throughout PBL. Note that this elementary school is one of the best schools in Beijing. Therefore, we might expect that more facilitation is required in other schools. However, the results showed that most students could recognize the situation where they did not need to search for any additional facts. Thus, students were not completely lost. We believe that it is practical to train students to recognize and focus on the relevant information when necessary. As we have already implemented a system that is able to detect different types of reading behaviors, one of the next steps is adding intervention to persuade students to keep their focus on situation-related information to reduce their cognitive load. Hopefully, this kind of training can help students solve these types of ill-structured problems more efficiently and successfully complete their PBL activities.

Another explanation for the students’ poor performance is that they do not perform well in discontinuous reading, as shown in the results reported by the 2009 PISA (OECD, 2010). Although the assessment results were from several years ago, fundamental changes have not been made in the Chinese reading instruction system. Most of the reading practices, especially those in elementary schools, are still well structured so that students can easily digest the learning contents. Although the test items in this assessment task clearly state what needs to be done, the given contents are distributed in different locations and mixed with other irrelevant information, which makes the assessment task require a more discontinuous reading ability instead of continuous reading ability.

The results suggest that many students are not good at acquiring factual knowledge by themselves to solve real-life problems. Therefore, teachers should be very careful in conducting PBL in elementary schools. Teachers are expected to provide a great amount of facilitation to their students. This paper does not try to prevent other teachers from using PBL in elementary schools, but instead remind them of the related issues. Notably, previous studies have shown that after students become comfortable with PBL, they no longer need too much facilitation (Holliday, 2006; Jalani & Lai, 2015).

In addition to students’ problem-solving outcomes, we also looked at how much invalid work they performed. Students who paid too much attention to irrelevant information tend to perform more invalid work. These students simply tried to work out every piece of information without thinking carefully. They might try their best to solve the problems but just not solve the problems using the correct method. It is possible that these students

<table>
<thead>
<tr>
<th>Reading behaviors</th>
<th>Short-reading strong relevant materials</th>
<th>Short-reading medium relevant materials</th>
<th>Short-reading weak relevant materials</th>
<th>Non-short-reading strong relevant materials</th>
<th>Non-short-reading medium relevant materials</th>
<th>Non-short-reading weak relevant materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation with task</td>
<td>-0.183</td>
<td>-0.198</td>
<td>-0.298</td>
<td>0.117</td>
<td>0.097</td>
<td>-0.417</td>
</tr>
<tr>
<td>performance</td>
<td>( p = .317 )</td>
<td>( p = .279 )</td>
<td>( p = .091 )</td>
<td>( p = .523 )</td>
<td>( p = .597 )</td>
<td>( p = .018 )</td>
</tr>
<tr>
<td>Correlation with the</td>
<td>0.156</td>
<td>0.153</td>
<td>0.398*</td>
<td>0.102</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>sign of shallow</td>
<td>( p = .367 )</td>
<td>( p = .404 )</td>
<td>( p = .024 )</td>
<td>( p = .580 )</td>
<td>( p = .934 )</td>
<td>( p = .937 )</td>
</tr>
</tbody>
</table>

Note. \( p < .05 \).

Discussion

<table>
<thead>
<tr>
<th></th>
<th>Short-reading strong relevant materials</th>
<th>Short-reading medium relevant materials</th>
<th>Short-reading weak relevant materials</th>
<th>Non-short-reading strong relevant materials</th>
<th>Non-short-reading medium relevant materials</th>
<th>Non-short-reading weak relevant materials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.183</td>
<td>-0.198</td>
<td>-0.298</td>
<td>0.117</td>
<td>0.097</td>
<td>-0.417</td>
</tr>
<tr>
<td></td>
<td>( p = .317 )</td>
<td>( p = .279 )</td>
<td>( p = .091 )</td>
<td>( p = .523 )</td>
<td>( p = .597 )</td>
<td>( p = .018 )</td>
</tr>
<tr>
<td></td>
<td>0.156</td>
<td>0.153</td>
<td>0.398*</td>
<td>0.102</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>( p = .367 )</td>
<td>( p = .404 )</td>
<td>( p = .024 )</td>
<td>( p = .580 )</td>
<td>( p = .934 )</td>
<td>( p = .937 )</td>
</tr>
</tbody>
</table>

Note. \( p < .05 \).
may study very hard in PBL, but still not learn the appropriate lessons from the task. Making students realize and focus on what they need to know seems straightforward, but is not easy to implement in practice. Zhang et al. (2014) found that high school students also tend to simply work out every piece of work whether or not it is useful for solving their problem.

After identifying these issues, how can we best help the students? First of all, our assessment system can be modified to train students to self-regulate their problem-solving process and reduce their cognitive loads by giving them hints to avoid placing too much of their attention on irrelevant materials. Secondly, schools should urge their teachers to encourage their students to practice more discontinuous reading before doing PBL to train their ability to synthesize information.

In PBL, students often need to seriously think about what they need to know. While making these decisions, students have to relate the current situation and information to their prior knowledge, which is considered as a deep approach to learning (Crooks & Alibali, 2013; Dolmans et al., 2015). The design of our system makes it possible to detect the moment a student decides what is necessary to know. If PBL’s learning materials can be arranged according to our system, i.e., put all the learning materials into the materials center and label each document, the system may also help students during the PBL process.

Conclusion and limitations

This paper introduced a technology-enhanced problem-solving ability assessment system that can track how students solve problems by analyzing their reading behaviors. In our experiment, most of the fifth-grade students were able to recognize whether to search relevant information for solving a specific task, but often failed in locating the relevant information when needed. The results also suggested that the distractions of irrelevant information could lead to bad problem-solving performance as well as shallow thinking.

The study clearly has several limitations. Firstly, the sample size is relatively small, and the participants are from the same class. Readers should be careful while generalizing our conclusion into a bigger population. Secondly, this study focused on exploring how well students identified text information to solve problems. Students might perform differently when the information is presented as figures, tables, or other formats. Thirdly, the study cannot externalize how students process the identified information in their minds. This is an aspect of our future research.

Acknowledgments

This work was supported in part by China Postdoctoral Science Foundation under projects number 2017M610054, and by Philosophy and Social Sciences Research of Chinese Ministry of Education under projects number 16JZD043.

References


An Eye Tracking Study of High- and Low-Performing Students in Solving Interactive and Analytical Problems

Yiling Hu, Bian Wu and Xiaqing Gu*
Department of Educational Information Technology, East China Normal University, Shanghai, China // ylhu@deit.ecnu.edu.cn // bwu@deit.ecnu.edu.cn // xqgu@ses.ecnu.edu.cn

*Corresponding author

ABSTRACT
Test results from the Program for International Student Assessment (PISA) reveal that Shanghai students performed less well in solving interactive problems (those that require uncovering necessary information) than in solving analytical problems (those having all information disclosed at the outset). Accordingly, this study investigates information-processing strategies in solving these two types of complex problems. High- and low-performing students’ eye fixations within certain areas of interest (AOIs) as well as their visits between AOIs were recorded as they solved problems on a computer. High-performing students had long fixation durations for the analytical problem and the problem-solving stage of the interactive problem. With regard to the problem exploration stage of the interactive problem, high-performing students had short fixation durations. These findings demonstrate how students had varying information-processing strategies for different types of problem-solving scenarios at different problem-solving stages. The implications of this study are discussed.

Keywords
Analytical problem solving, Interactive problem solving, Eye tracking, Strategic thinking, Reasoning

Introduction
Problem solving is a critical competence in daily life and work. The development of problem-solving competence has drawn increasing attention, not only from within the domain of educational practice, but also from commentators in society more widely. The term “problem” suggests various qualifiers for its definition and differentiation: these include “well-structured,” “ill-structured,” “domain-dependent,” “domain-independent,” “static,” “dynamic,” “interactive,” “analytical,” “visual,” “creative,” and “design” (Knoblich, Ohlsson, & Raney, 2001; Greiff, & Fischer, 2013; Jonassen, 2014; van Meeuwen et al., 2014). While researchers have confirmed the occurrence of certain general competencies across all types of problem solving, distinctive competencies associated with solving certain types of problem have also been identified (Jonassen, 2010). Accordingly, learning to solve only certain types of problem will not guarantee the mastery of competencies necessary to solve others (van Merriënboer, Clark, & De Croock, 2002). Therefore, research on the cognitive processes involved in solving a range of problem types remains important for fostering the development of student problem-solving competencies.

PISA but also our own previous study (Hu, Wu, & Gu, 2017; OECD, 2014) asserted that Shanghai students performed less satisfactorily when solving interactive problems (i.e., those that require uncovering necessary information) compared to performance in solving analytical problems (i.e., those that have all information disclosed at the outset). The possibility that problems in formal school contexts and those in “real-world” contexts are fairly different in terms of structure as well as complexity (Jonassen, 2014) might explain students’ different performances in solving analytical, school-like, and interactive complex problems.

Cognitive information processes of problem solving
Mayer and Wittrock (2006) defined the problem-solving process in terms of transformations of a given state into a goal state when no evident method of solution is available. General problem-solving processes may involve various cognitive competencies. These include reasoning, critical thinking, decision making, argumentation, and metacognition (Jonassen, 2014). They also may involve different cognitive activities, such as problem exploration, knowledge representation, problem solution planning, and evaluation (Mayer, 2003). Two typical types of problem have generally been investigated (Fischer et al., 2015; Greiff, Wüstenberg, & Avvisati, 2015; Wirth, & Klieme, 2003) and then have been embedded in international large-scale assessment: one concerns analytical problems, and the other concerns interactive problems (OECD, 2014).

In analytical problem solving, the problem solver needs to structure, represent, and integrate information to construct a problem solution (Simon, 1975). Given the static nature of analytical problems, the problem-solving
process is fairly straightforward, with limited opportunity for updating the problem schema, such that mental effort is mainly invested in manipulating operations to change initial states toward some target state. By contrast, in interactive problem solving, the problem situation may change over time, such that changing one variable in a task may lead to manifold changes in other variables (Funke, 2010). A problem solver cannot only rely on the information given at the outset. They must work out a plan to search the space of information, propose and adapt his or her hypotheses, as well as executing the plan to collect information to corroborate or refute the hypotheses and reach a final goal (Greiff & Fischer, 2013). Therefore, a problem solver should keep restructuring or updating the problem representation in order to solve incongruence or to integrate new information: they must adjust problem exploration and reasoning strategies accordingly. Moreover, the problem-solving pathway may vary from one problem solver to another because of the interrelationship between dynamic problem exploration and problem solving (Sprenger & Dougherty, 2012).

Regarding the general information processes during problem solving, several typical strategies have been widely examined. For example, during the problem exploration stage, one strategy has been termed “cue-wise” or “history-cue,” in which people use previously identified cues to drive further problem space exploration (Renkewitz & Jahn, 2012; Sweller, Mawer, & Howe, 1982). Another strategy is termed “trial-and-error,” in which random attempts are made to select the operant. While during the problem-solving process, one strategy has been called “schema-driven,” in which individuals conduct forward reasoning on the basis of the constructed knowledge schema (Gick, 1986). A further strategy has been called “means-end,” in which attempts are made to reduce the difference between the current state and the goal state.

The assertion of Fischer et al. (2015) that analytical problem-solving and interactive problem-solving address a common core of problem-solving competence cannot explain the reason why Chinese students perform better in the former type of problem than they do in the latter (Hu et al., 2017; OECD, 2014). Certain studies have confirmed that solving interactive and analytical problems requires different competencies in terms of both knowledge acquisition and knowledge application. For example, Greiff and Fischer (2013) argued that problem exploration as well as representation each require strategic knowledge on capturing critical information, constructing well-organized problem schema, and generating, as well as testing, hypotheses. Wirth and Klieme (2003) asserted that interactive problem solving includes aspects of self-regulated learning, as well as processes of feedback management.

**Eye tracking in problem solving**

Eye-tracking technology offers an attractive and powerful tool to reveal the cognitive processes of student learning (Lai et al., 2013). Many studies use eye-tracking techniques for focusing on different learning domains, such as multimedia learning (van Gog & Scheiter, 2010), conceptual learning (Ariasi & Mason, 2011), language learning (Whitford & Titone, 2012), and category learning (Rehder & Hoffman, 2005). Given that a complex problem-solving process requires the integration of different competencies, multimodal data (such as the log data of learning behavior (Worsley & Bilkstein, 2015)), eye movements, learning outcomes, and self-report, can all provide distinct analytic perspectives for deepening our understanding of this field. For example, visual attention during problem-solving, as measured by eye movements, can be used to reveal the critical aspects of the process that traditional measures, such as solution time and accuracy or behavior-level keystrokes, cannot address (Grant & Spivey, 2003; Lee & Anderson, 2001; Tsai, Hou, Lai, Liu, & Yang, 2012; van Meeuwen et al., 2014; Yeh, Tsai, Hsu, & Lin, 2014).

For the recording of eye movements, two basic measures are used: first eye fixation, which is a relatively stable state of eye position and, second, saccades, which are the rapid eye movements between eye fixations (Rehder, Colner, & Hoffman, 2009). Based on these two measures, eye-tracking indices of three forms can be used for analysis. Counts are also possible, both spatial and temporal, such as number of fixations, fixation positions, and saccade durations (Radach & Kennedy, 2004). In addition, researchers may be interested in identifying the information that learners gaze at, typically called “the area of interest” (AOI), but also the gazing duration, and the gazing sequence. In this way, eye-tracking indices can provide rich information in relation to the different AOIs and transitions among different AOIs. However, how to interpret measurements effectively in various learning contexts is a vital issue when applying the eye-tracking method in educational research (Lai et al., 2013). For example, mental effort investment can be directly measured using temporal scales (such as fixation duration) (Saß, Schütte, & Lindner, 2017); count measures of fixation can be used to reveal the importance of visual materials (Balslev et al., 2012); and different patterns of saccadic eye movement may be taken to imply different information-processing strategies (van Meeuwen et al., 2014). Lai et al. (2013) has reported that previous studies employed temporal indicators most frequently, followed by spatial and count indicators.
By examining major themes in eye-tracking studies, Lai et al. (2013) argued that considerable effort should be focused on the process of reasoning. For example, Renkewitz and Jahn (2012) investigated memory retrieval during decision making and identified different gaze patterns associated with different underlying memory retrieval strategies. Thibaut and French (2016) explored search strategies reflected in different looking patterns pertaining to the analogy problem. However, most tasks in these studies lasted for only seconds and so cannot be considered complex problems. In other words, a body of knowledge on information processing strategies in complex problem-solving performances has not yet been realized.

Research hypotheses

The general assumption underlying our study is that information-processing strategies during problem solving will be reflected in eye movement patterns. Moreover, that eye movements will be aligned with different problem-solving performances and relate to efficiencies. Specifically, we propose the following research questions:

Research question 1: How does eye movement differ between high- and low-performing students in solving analytical and interactive problems?

Research question 2: What different information-processing strategies can be inferred from the eye movements of high- and low-performing students during the solving of analytical and interactive problems?

Methods

Participants

Twenty-eight undergraduate students from the first year \((n = 12)\), the second year \((n = 10)\) and the third year \((n = 6)\) of a university in Shanghai voluntarily took part in this study. There were no restrictions relating to the discipline and prior knowledge of the participants. The only requirement was that all participants should have normal eye sight or wear glasses, because the computer screen would be placed approximately 80 cm from the eyes of the participants to collect eye movement data. Ten participants (35.7%) were male. Their average age was 20.7 years with age ranging from 19 to 21. Every participant received a small souvenir for their participation.

The reason for selecting undergraduates instead of K-12 students is twofold. First, given that finding K-12 students to participate in the study in our laboratory was difficult, we recruited undergraduate students from our own university on the basis of convenience sampling. Second, our purpose was to investigate information-processing strategies between high- and low-performing students in solving two types of problems, rather than to identify the problem-solving features of Chinese K-12 students. Therefore, as we stress in the Discussion section, although this study may elucidate possible explanations of Chinese students' performance in PISA tests, the present findings may not be generalized to other levels of students.

Materials

We adopted two PISA test problems and built a computer-based problem-solving environment for their presentation. One was a rule induction problem, called 'birthday party', which belongs to the analytical problem category. This problem required students to arrange seven different animals using drag and drop methods around a table to meet all conditions in a given list. A total of nine conditions were present in the study. For example, one condition was that the tiger sits neither next to the elephant nor next to the mouse. The score of this problem task ranges from zero to seven depending on how many animals were correctly arranged.

The other problem was a multi-variable system problem, which is a type of interactive problem. In this problem, a computer-simulated forest system was presented, which included three manipulation variables (inputs) to affect three quantities (outputs): temperature, precipitation, and population of species. The problem involved two stages, i.e., problem exploration and problem solving. The rationale of this two-stage interactive problem was for students to explore the problem situation as well as to understand the underlying system (system representation) and then to apply obtained knowledge (represented system) to solve a new problem situation.
In the first stage, the students were asked to figure out the relations between input variables and output variables through adjusting the input values by clicking a button to simulate the system to the next time step, as well as by reading the line charts of output values ranging from the very beginning until the present. The students then draw lines between inputs and outputs to represent the relations of this multivariable system. Given that this multivariable system contains four relations between inputs and outputs, the score for this problem task ranges from zero to four, with each correct relation getting one point.

In the second stage, the correct relationships underlying this multi-variable system were provided along with current and target output values, regardless of whether students identify the relations correctly in the first stage. The students were asked to determine the most efficient way to manipulate input variables to reach the output targets. In other words, the fewer simulation steps, the better. The total score is four, with each of three output goals being reached getting one point and with the completion of the task within 7 steps getting one point.

Research procedure and apparatus

The study took place in our laboratory, where the participants conducted the experiment one at a time. All participants completed the tasks in three steps. First, the test procedure was described, including the use of the problem-solving environment and the eye tracking system. Second, the participants were seated at an approximate distance of 80 cm in front of a 17-inch screen with a 1280 x 1024 pixel resolution. They submitted to eye-tracking calibration and then addressed the demonstration problem in this computer-based problem-solving context. Third, they were required to solve one analytical problem and one interactive problem. The whole experiment lasted approximately 30 minutes. Data were collected using the Tobii T120 eye tracker (120 Hz sampling rate) manufactured by Tobii Technology (Stockholm, Sweden). This eye tracker is not intrusive. A large “track box” provides the user with the freedom of head movement and considerable comfort. Eye movement data during problem solving were screen-captured and analyzed with Tobii-Studio (3.1.6) software.

Data collection and analysis

Problem-solving performance data were collected automatically by the system in a log file, which included the scores of problem-solving solutions and the test duration for solving each problem. For the analysis of eye movement patterns, the AOI was defined first. One rectangle area called the problem information area (PIA) and one called the problem operation area (POA) were selected in the three problem tasks (i.e., analytical problem, the problem exploration stage of the interactive problem, and the problem-solving stage of the interactive problem). The POA in analytical and interactive problems is the area where problem solvers select operations to solve the problem. On the other hand, PIA in the analytical as well as the interactive problem is the area where all the information on the problem is presented at the outset and where simulated dynamic information is also presented. In addition, one rectangular area called the problem representation area (PRA) was selected in interactive problems, where problem solvers were required to represent their understanding of the interactive problem and also apply this problem representation to solve a new problem situation. Figures 1 to 3 depict the defined AOIs.
Eye movement indicators generated from the Tobii studio included fixation duration, fixation count, visit duration, and visit count on all target AOIs. Fixation in this study was identified as a gaze point that lasted for at least 60 ms, which is the minimum time for a stimulus to travel from the retina to the region of the brain that processes the information (Chien et al., 2015; Rayner, 1998). Fixation duration was calculated as the average length of fixations in a look zone. Fixation count was the number of fixations at a certain area on the screen. Visit time refers to the period from entering a specific AOI until moving out of this area and is calculated as the durations between the first and the last consecutive fixations in the AOI.

**Figure 2.** Eye movements in the problem exploring stage of interactive problems—(a) eye fixations and saccades of a high-performing student, (b) eye fixations and saccades of a low-performing student. *Notes.* POA stands for problem operation area; PIA stands for problem information area; PRA stands for problem representation area.

**Figure 3.** Eye movements in the problem-solving stage of the interactive problem—(a) eye fixations and saccades of a high-performing student, (b) eye fixations and saccades of a low-performing student. *Notes.* POA stands for problem operation area; PIA stands for problem information area; PRA stands for problem representation area.

We used the mean test performance in each problem-solving test as a threshold to divide the participants into the low- and high-performing groups, assuming that performance for all participants may not be consistent across all three problem tasks. We conducted the Chi-squared test of group division in each problem task and confirmed that group divisions in all three problem tasks differed from one another: (1, N = 28) = 5.038, p = .051 for tasks 1 and 2; (1, N = 28) = 2.798, p = .125 for tasks 1 and 3; (1, N = 28) = 1.152, p = .433 for tasks 2 and 3.

First, descriptive statistics and independent sample *t*-tests were conducted to examine the difference of eye movement patterns between the high- and the low-performing groups on analytical and interactive problem-solving processes. Second, purposeful sampling was used to select two students for case study, considering the typicality of the case. One student from the high-performing group and one from the low-performing group were selected to analyze their eye movement sequences qualitatively.
Results

Comparison of eye movement difference between the high- and low-performing groups

Analytical problem solving

To test the eye movement difference between the two groups in the analytical problem context, independent sample t-tests were conducted (see Table 1). Regarding PIA, the students in the high-performing group had significantly longer fixation durations compared with the students in the low-performing group. Significantly fewer fixation counts are found for the high-performing group than for the low-performing group. Regarding the completion time of solving the analytical problems, significantly shorter test durations were found for the high-performing group than for the low-performing group.

Table 1. Eye movement differences between the high- and low-performing groups in analytical problem solving

<table>
<thead>
<tr>
<th>AOI</th>
<th>Indicator</th>
<th>Low (N = 11)</th>
<th>High (N = 17)</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Problem</td>
<td>Fixation duration(s)</td>
<td>.26</td>
<td>.11</td>
<td>.45</td>
</tr>
<tr>
<td>information area (PIA)</td>
<td>Fixation count</td>
<td>163.69</td>
<td>60.47</td>
<td>74.60</td>
</tr>
<tr>
<td></td>
<td>Visit time(s)</td>
<td>2.56</td>
<td>.95</td>
<td>2.80</td>
</tr>
<tr>
<td>Problem</td>
<td>Fixation duration(s)</td>
<td>.44</td>
<td>.16</td>
<td>.43</td>
</tr>
<tr>
<td>operation area (POA)</td>
<td>Fixation count</td>
<td>193.85</td>
<td>198.91</td>
<td>230.00</td>
</tr>
<tr>
<td></td>
<td>Visit time(s)</td>
<td>2.26</td>
<td>.75</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>Visit count</td>
<td>38.46</td>
<td>30.03</td>
<td>35.67</td>
</tr>
<tr>
<td>Test duration(s)</td>
<td></td>
<td>237.10</td>
<td>83.80</td>
<td>137.03</td>
</tr>
</tbody>
</table>

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

Problem exploring stage of interactive problems

To test the eye movement difference between these two groups in the problem-exploring stage of the interactive problem, we conducted independent sample t-tests (see Table 2). Regarding POA, significantly shorter fixation durations were found for the high-performing group than for the low-performing group and significantly fewer visit counts for the high-performing group than for the low-performing group. Regarding PIA, significantly shorter fixation durations were found for the high-performing group than for the low-performing group and significantly fewer visit counts for the high-performing group than for the low-performing group. Regarding the completion time of the problem exploration stage of the interactive problem solving, a significantly shorter test duration was found for the high-performing group than for the low-performing group.

Table 2. Eye movement differences between the high- and low-performing groups in the problem exploration stage of interactive problem solving

<table>
<thead>
<tr>
<th>AOI</th>
<th>Indicator</th>
<th>Low (N = 13)</th>
<th>High (N = 15)</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Problem</td>
<td>Fixation duration(s)</td>
<td>.62</td>
<td>.21</td>
<td>.32</td>
</tr>
<tr>
<td>operation area (POA)</td>
<td>Fixation count</td>
<td>36.85</td>
<td>21.47</td>
<td>38.80</td>
</tr>
<tr>
<td></td>
<td>Visit time(s)</td>
<td>1.51</td>
<td>.64</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>Visit count</td>
<td>17.54</td>
<td>6.63</td>
<td>7.47</td>
</tr>
<tr>
<td>Problem</td>
<td>Fixation duration(s)</td>
<td>.58</td>
<td>.22</td>
<td>.33</td>
</tr>
<tr>
<td>information area (PIA)</td>
<td>Fixation count</td>
<td>56.77</td>
<td>36.37</td>
<td>38.73</td>
</tr>
<tr>
<td></td>
<td>Visit time(s)</td>
<td>1.77</td>
<td>.47</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>Visit count</td>
<td>14.31</td>
<td>4.29</td>
<td>8.27</td>
</tr>
<tr>
<td>Problem</td>
<td>Fixation duration(s)</td>
<td>.41</td>
<td>.14</td>
<td>.40</td>
</tr>
<tr>
<td>representation area (PRA)</td>
<td>Fixation count</td>
<td>3.08</td>
<td>1.71</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td>Visit time(s)</td>
<td>.23</td>
<td>.26</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>Visit count</td>
<td>4.69</td>
<td>3.99</td>
<td>5.20</td>
</tr>
<tr>
<td>Test duration(s)</td>
<td></td>
<td>89.15</td>
<td>32.88</td>
<td>51.09</td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01.
Problem-solving stage of interactive problems

To test the eye movement difference between these two groups in the problem-solving stage of the interactive problem, we conducted independent sample t-tests (see Table 3). Regarding POA, significantly longer fixation durations were found for the high-performing group than for the low-performing group and significantly fewer visit counts for the high-performing group than for the low-performing group. Regarding PIA, significantly longer fixation durations were found for the high-performing group than for the low-performing group and fewer visit counts for the high-performing group than for the low-performing group. Regarding PRA, significantly longer fixation durations were found for the high-performing group than for the low-performing group. Regarding the completion time of the problem-solving stage of interactive problem solving, a significantly shorter test duration was found for the high-performing group than for the low-performing group.

Table 3. Eye movement differences between the high- and low-performing groups in the problem-solving stage of interactive problem solving

<table>
<thead>
<tr>
<th>AOI</th>
<th>Indicator</th>
<th>Low (N = 10)</th>
<th>High (N = 18)</th>
<th>t-test</th>
<th>MD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem operation</td>
<td>Fixation duration(s)</td>
<td>.27 .14</td>
<td>.53 .25</td>
<td>−12.28</td>
<td>−.86</td>
<td>.40</td>
</tr>
<tr>
<td></td>
<td>Fixation count</td>
<td>74.00 49.69</td>
<td>61.72 26.61</td>
<td>−15.98</td>
<td>−3.33</td>
<td>.01*</td>
</tr>
<tr>
<td>Problem information</td>
<td>Visit time(s)</td>
<td>1.07 .35</td>
<td>1.11 .36</td>
<td>.04</td>
<td>.29</td>
<td>.77</td>
</tr>
<tr>
<td></td>
<td>Visit count</td>
<td>37.20 14.45</td>
<td>21.22 6.25</td>
<td>−15.98</td>
<td>−3.33</td>
<td>.01*</td>
</tr>
<tr>
<td>Problem representation</td>
<td>Fixation duration(s)</td>
<td>.29 .12</td>
<td>.52 .20</td>
<td>−15.99</td>
<td>−9.5</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>Fixation count</td>
<td>84.60 50.52</td>
<td>68.61 22.71</td>
<td>−10.77</td>
<td>−4.00</td>
<td>.00**</td>
</tr>
<tr>
<td>Problem representation</td>
<td>Visit time(s)</td>
<td>1.25 .40</td>
<td>1.41 .56</td>
<td>.16</td>
<td>.79</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>Visit count</td>
<td>27.60 9.26</td>
<td>16.83 5.07</td>
<td>−10.77</td>
<td>−4.00</td>
<td>.00**</td>
</tr>
<tr>
<td>Test duration(s)</td>
<td></td>
<td>130.54 67.97</td>
<td>69.41 22.34</td>
<td>−61.13</td>
<td>−2.76</td>
<td>.02*</td>
</tr>
</tbody>
</table>

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

Case analysis of eye movement patterns

The complete eye movement record in solving one problem lasted several minutes and contained a significant number of circles (fixations) and lines (saccades) that make the figures difficult to read. Therefore, this preliminary study presented eye movement within a small time window (around ten seconds) to illustrate the information-processing strategies during the problem-solving stage reflected from the eye movement patterns.

Analytical problem-solving strategies

Analytical problem solving requires planning and execution processes, in which rule search strategy and reasoning are needed. When examining the fine-grained fixation transitions of problem solving, we can see clearly how two strategies were conducted by a high-performing student and a low-performing student, respectively. For example, after the application of the first three rules, the current problem state was four animals, including monkey, horse, dog, and tiger, settled on the table. A high-performing student (see Figure 1(a)) gazed at rule four, which said that the elephant sits next to either the turtle or the pig (includes none of those four animals), and then saccade to the next one until he or she gazed at rule six (includes a cue about tiger that is already on the table). Subsequently, the application of rule six reduced the options of choice and moved the problem state a step close toward the target. By contrast, a low-performing student (see Figure 1(b)) visited the problem-solving area after gazing at each rule, not noticing that the application of certain rules (e.g., rule 4 and rule 5), in fact, caused additional alternatives and expanded the problem space.

Searching strategies in the problem exploration stage of interactive problems

In the problem exploration stage of a multivariable system, the control of variables strategy (CVS) is an ideal strategy and is generally defined as a learning goal for a novice problem solver to grasp (Greiff, Wüstenberg, & Avvisati, 2015; Lazonder, 2014). The CVS refers to varying one thing at a time to identify the relations between
multi-inputs and multi-outputs. The interactive behavior of a CVS adopter (see Figure 2(a)) shows the manipulation of one variable at a time (i.e., fixation transition between POA and PIA after gazing at one input variable), which cannot be found in the behavior of a non-CVS adopter (i.e., fixation transition within POA among different input variables before transiting back and forth between POA and PIA) (see Figure 2(b)).

**Reasoning strategies in the problem-solving stage of interactive problems**

In the problem-solving stage of interactive problems, the purpose is to manipulate three inputs and simulate the system to reach three target output values in the fewest steps. Two typical strategies, i.e., schema-driven and means-end, are commonly mentioned in the problem-solving domain (Patel, Groen, & Frederiksen, 1986; van Meeuwen et al., 2014). As described by van Meeuwen et al. (2014), the use of means-end analysis would be manifested by frequently focusing on the goal to reason backward, which is an effort-demanding strategy, whereas the use of schema-driven strategies would be manifested by frequently focusing on the elements that are affected by the problem-solving steps to reason forward, which is a highly efficient strategy. Figure 3(a) illustrates that a high-performing student gazed at PRA and PIA before transiting fixation within POA among different input variables (to work out a solution and conduct a schema-driven strategy) and with little back and forth between POA and PIA. While a low-performing student (see Figure 3(b)) had frequent fixation transitions between POA and PIA (to conduct a means-end approach) with a few fixations on PRA.

**Discussion**

**Analytical problem-solving**

In analytical problem solving, two strategies were identified in relation to different performances. High-performing students adopted a heuristic search strategy to narrow the problem space efficiently. They searched from the rule list to identify the rule that contained the critical information relating to the current problem state. However, low-performing students adopted trial-and-error approaches to apply the rules in the original sequence as presented in the problem statement.

The aforementioned descriptions of the two strategies cannot be easily identified from the behavior during problem solving. However, according to the eye movement indicators of different AOIs, we find distinctive eye movement patterns in line with these two strategies. Therefore, compared to low performing students, high-performing students have longer fixations and fewer visit counts in PIA, which, in turn, causes shorter test durations. In analytical problem-solving, the cue-wise strategy of relevant rule selection costs extra time in gazing at PIA that the trial-and-error strategy needs, whereas the latter strategy leads to frequent back and forth between PIA and POA due to the expansion of the problem space. Therefore, we can infer from these critical indicators, including fixation durations and visit counts, whether or not the problem solver adopted a superior strategy. Goldberg and Kotval (1999) also argued that a few fixation counts indicate highly efficient search.

Consistent with Renkewitz and Jahn’s study (2012), the case analysis of eye movement patterns further verified that the fixation duration of cue-wise strategy adopters’ on rules—including specific cues in the current solution state—reflected the rule selection order by this strategy. On the other hand, trial-and-error strategy adopters had similar fixation durations on all rules. The quantitative analysis of eye-tracking indicators and the qualitative case analysis of eye movement patterns corroborate the argument that information reduction abilities can be demonstrated by long fixation duration and a few visit counts (van Meeuwen et al., 2014).

**Strategic planning and reasoning in interactive problem-solving**

In interactive problem solving for a multi-variable system problem, two consecutive stages were addressed: problem exploration and problem solving. Different information-processing strategies in these two stages were confirmed and were in line with previous studies (Saß, Schütte, & Lindner, 2017). For the first stage, the present study identified information-processing strategies that reflect the adoption and non-adoption of CVS, which is further confirmed by two distinctive eye movement patterns (Greiff et al., 2015). As for eye movement patterns, CVS adopters showed shorter fixation durations in POA and PIA compared with non-CVS adopters. Moreover, low-performing non-CVS adopters shifted their fixation back and forth among different AOIs frequently (i.e., additional visit counts), thereby suggesting that in the knowledge representation phase these students failed to
build up the correct mental model that would be required to constantly re-evaluate their mental model with incoming information to solve incongruence.

For the second stage, we found two types of reasoning strategy. The first type is schema-driven forward reasoning that leads to highly efficient and effective performances (Patel, Arocha, & Zhang, 2005; van Meeuwen et al., 2014; Wills, Lavric, Croft, & Hodgson, 2007). The second type is means-end backward reasoning that is an effort-demanding strategy. Students adopting this strategy search the problem space continuously to reduce the difference between the current state and the goal state (Simon, 1975). Our findings corroborated that low-performing students had high visit counts and short fixation durations, which suggests that their attention was focused on frequent adjustment of input variables and the comparison of the current output values as well as the target values that a means-end strategy would exhibit. By contrast, high-performing students tend to retrieve and apply their previously constructed problem schema to solve current problems with long fixation durations in PRA, POA, and PIA. The short visit counts in PIA and POA also suggest the well-planned reasoning strategy of high-performing students. Further case analysis of two typical examples triangulates the quantitative findings of different strategy adoptions. These findings are consistent with Cook et al. (2008), which indicates that students with high prior knowledge had longer fixation durations on critical information compared with those with low prior knowledge.

Comparison of information-processing strategies between analytical and interactive problem solving processes

By comparing the eye movement dynamics between analytical and interactive problem solving processes, we found similar fixation duration patterns in analytical problems and the problem-solving stage of interactive problems. Our findings add new knowledge to the work of van Meeuwen et al. (2014), by suggesting that not only the efficient scan path of relevant information but also the capability of retrieving a problem schema, as identified during problem exploration, can be predicated by long fixation durations within a certain AOI.

Contrary to analytical problem solving, high-performing students in the problem exploration stage of interactive problem solving showed short fixation durations, which may suggest that restructuring dynamic information during an interactive problem-solving process is the end result of active memory search processes that differ from cognitive processes underlying the analytical problem-solving process (Fleck, 2008). Analytical problem solving requires cue-wise rule induction, (i.e., long fixation durations within the problem information area), whereas interactive problem solving requires CVS strategies for problem exploration (few visit counts) followed by knowledge-schema-driven problem solving (long fixation duration within a problem representation area). Therefore, this study adds empirical findings to Chen et al. (2014) and further argues that different problem-solving performances can be predicted by different eye movement indicators according to different problem types.

Conclusion and limitations

The findings of this exploratory study verify the following. (1) In analytical problem solving, high-performing students had longer fixation durations as well as fewer fixation counts and visit counts in PIA compared with low-performing students, while the high- as well as low-performing groups evidently adopt the cue-wise strategy and the trial-and-error strategy, respectively. (2) In the problem exploration stage of interactive problems, high-performing students had short fixation durations and a few fixation counts in PIA and POA, and the high-performing group evidently adopts the CVS strategy. (3) In the problem-solving stage of interactive problems, high-performing students had long fixation durations in PIA, POA, and PRA as well as a few visit counts in PIA and POA, and the high- as well as low-performing groups evidently adopt the schema-driven strategy and the means-end strategy, respectively.

The study also validates the notion that different problem tasks may require different visual attention and mental effort allocations in different parts of a problem. For example, in analytical problem solving, PIA is more important compared with POA for critical information selection; in the problem exploration stage of interactive problems, considerable attention should be given to the relation between PIA and POA; while in the problem-solving stage of interactive problems, all three areas—including PRA, PIA, and POA—should draw considerable attention and in the correct sequence to reduce fixation transitions among different areas.
To foster the deep learning of students during problem solving, we need initially to extract observable and critical indicators (Knoblich, Ohlsson, & Raney, 2011; van Meeuwen et al., 2014). Only then can we possibly provide visual cues as guidance or through personal feedback during problem-solving activities (Jarodzka et al., 2012). This study proves that, through examining fixation duration, fixation count, and visit count, we can tell why students show different levels of performance in different problem tasks, which is, in turn, a result of adopting different information processing strategies.

This study opened a new window for the future analysis of complex problem-solving competencies on the basis of eye movement data. Previous studies have corroborated that commonly used problem-solving analytic approaches, such as think-aloud and performance analyses, may involve certain side effects, including inaccurate data, missing data, and misleading data (such as arise from the issue of gaming the system (Baker et al., 2008)). Eye movement data, which are real-time, rich, and unobtrusive as well as involving eye-mind coordination, can provide fine-grained analysis to build a deep understanding of problem solving and to triangulate with evidence obtained from traditional ways of problem-solving assessment. In short, the results verify that an eye-tracking approach yields valuable insights on information processing strategies during problem solving. We found distinctive patterns of eye movements in solving analytical problems, as well as in exploring and solving interactive problems.

This study chose undergraduate students for the problem-solving test using PISA test problems that are appropriate for K-12 students. Although the purpose of this study was to investigate different information processing strategies adopted by high- and low-performing students to solve the two types of problems, the mismatch between student levels and problem complexity can be considered a limitation. So the findings of this study may apply to the same level of students but may not be generalized to other levels of students directly. The small sample size is another limit to the generalizability of the findings of this study. A further limitation is that only two samples were selected in each problem scenario to analyze fixation transition manually. In future studies, we must consider using sequential pattern recognition technology to analyze a large sample of fixation transitions automatically. A fourth limitation is concerned with the presence of only two problems in this study, one for the analytical problem and one for the interactive problem. Possibly, different types of analytical and interactive problems may be aligned with different information processing strategies. Therefore, whether these findings of eye movement patterns can be generalized to other interactive and analytical problem-solving processes remains an open question. Lastly, future studies need to consider cognitive load and further differentiate the cause of eye movement dynamics between the mental effort of knowledge retrieval and strategic thinking (Amadieu et al., 2015) because problem solving requires information-processing strategies and memory capacities.

Acknowledgements

This study is supported by the Humanity and Social Science Youth foundation from the Ministry of Education of China (grant number 16YJC880085) and the Peak Discipline Construction Project of Education from East China Normal University, China. The authors would like thank Jiayu Zhu for her generous help in data collection analysis. The authors also appreciate the thoughtful comments from Charles Crook on the draft version. The authors are also grateful to the editors and anonymous reviewers who provided us with valuable comments to improve the manuscript.

References


**Formative Assessment in Complex Problem-Solving Domains: The Emerging Role of Assessment Technologies**

Kaushal Kumar Bhagat1 and J. Michael Spector2

1Smart Learning Institute, Beijing Normal University, China // 2Department of Learning Technologies, University of North Texas, TX, USA // kkntmu@hotmail.com // mike.spector@unt.edu

*Corresponding author

**ABSTRACT**

Much of the focus on learning technologies has been on structuring innovative learning experiences and on managing distance and hybrid learning environments. This article focuses on the use of technology as an important formative assessment and feedback tool. The rationale for this focus is based on prior research findings that suggest that timely and informative feedback promotes learning. The general purpose of this article is to promote a focus on formative assessment, especially with challenging problems so as to help develop critical thinking skills. This is not a meta-analysis although we encourage such studies so as to emphasis the role that formative assessment plays in supporting learning. Much of the prior research on formative assessment has not involved advanced digital technologies and very little of that research has focused on complex and challenging problem-solving tasks. We review prior work on the use of problem conceptualizations elicited during problem solving activities and on stealth assessments of learner choices and decisions in online activities. We present a conceptual framework based on prior research and theory for conducting formative assessments in real-time with regard to complex problem-solving tasks. We then present an elaboration of how formative assessment can be used to support learning a common intellectual skill involving a discrimination task and to develop an appropriate cognitive strategy for that kind of problem. We conclude with recommendations for further research on the use of technology in support of formative assessment.

**Keywords**

Cognitive strategies, Complex learning, Dynamic feedback, Formative assessment, Intellectual skills, Self-regulation

**Introduction**

There are three major findings from research on learning in the last 50 or more years (Narciss, 2008; Spector & Yuen, 2016): (a) time on task predicts learning outcomes, (b) formative feedback tends to improve learning, and (c) prior knowledge and experience influences learning. These three findings from learning research are related and have direct implications for formative assessment. First, however, we need to define the scope and purpose of formative assessment. Formative assessment is feedback provided to the learner during an instructional sequence or learning activity that is aimed at helping the learner succeed. Timely and informative feedback is essential for formative assessment to be effective, although the amount and timing of feedback should be appropriate for a particular learner. This is how prior learning and experience is related to formative assessment. More advanced learners require less support and may regard too much feedback as intrusive according to the research on cognitive apprenticeship (Collins, Brown, & Newman, 1987). In addition, learners who spend more time on a learning activity or task are likely to gain more understanding and competence, so formative feedback that encourages continued engagement is also likely to be more effective in support of learning (Spector & Park, in press).

In general, there is sufficient research on formative assessment to support learning simple tasks in well-structured domains with outcomes targeted at simple concepts and procedures. The explosion of new technologies makes such support ever more effective. What is less well understand is how best to support learning complex and ill-structured tasks and how best to use new technologies to support formative assessment in those situations. That gap is the focus of this contribution. The good news is that there are new techniques and findings emerging that should support progress in this important domain.

**Formative assessment research**

Research on formative assessment presents particular challenges, even though the benefits are widely acknowledged (Black & Wiliam, 1998; Dunn & Mulvonen, 2009; Fuchs & Fuchs, 1986). One challenge is to determine the influence of formative assessment on learning. If learning is defined as stable and persistent changes in what a person knows or can do (Spector & Yuen, 2016), then there is limited research as there are few studies that examine the persistence of learning and fading of knowledge and skills over time. A second
challenge involves how and when formative assessments are provided. That is to say, while a particular feedback may be intended as constructive and encouraging, it may be perceived as discouraging or disheartening. In addition, a delayed feedback may not serve well to support continued engagement in a learning activity or instructional sequence. Finally, determining whether it was a formative assessment that led to an impact on learning rather than other factors is rarely explored in a controlled manner.

One reason for the lack of randomized controlled studies involving formative assessment is an ethical about potentially disadvantaging a group of learners not being provided a preferred form of feedback. One way to address that concern is through a within subject design using repeated and alternating treatments. Such a study should be sufficiently long with sufficient alterations to potentially observe differences due to different forms of feedback and should also involve measures before, during, after and long after the instructional sequence. In such a study there is the possibility that a learner will begin to generate a form of self-assessment that is similar to the kind of formative feedback expected to be most productive in terms of learning gains. Such a meta-cognitive learning outcome can be determined and is a desirable outcome aligned with the notion of self-regulated learning (Butler & Winne, 1995).

Complex learning

Some learning tasks and problems are more complex than others. Complexity can result from there being many interacting factors, from some of those interactions being non-linear with delayed effects, from changes in the problem situation changing while a solution approach is being formulated, and from lack of learner familiarity with the general nature of the problem or problem domain (Spector, 2015; van Merriënboer, 2012). Such problems and learning tasks often have multiple acceptable solution approaches and solutions, which presents a unique challenge for formative assessment. If supportive and potentially corrective feedback is not provided early in the process and in a meaningful manner, then a learner may develop misconceptions that are difficult to overcome later in a learning progression.

One fundamental theoretical foundation that has existed for some time is based on the notions of authentic learning introduced by Dewey (1938) and more recently elaborated in the form of situated learning (Lave & Wenger, 1991). These theoretical foundations strongly support two things. First, it is important to use meaningful and realistic problems to help develop complex problem-solving skills. Of course, prerequisite knowledge cannot be assumed, although the problem-based learning community argues that much prerequisite knowledge should be introduced while working on real problems (Barrows & Tamblyn, 1980). Complex problems that occur in so many domains are typically addressed by small groups of specialists. As a consequence, developing sufficient competence to be recognized as a contributing member of a problem-solving team in an important consideration (Lave & Wenger, 1991; Milrad, Spector, & Davidsen, 2003). In short, to develop complex problem-solving skills involves the development of competence and confidence. Previous attempts that emphasis summative assessments have tended to focus on the competence aspect of that developmental process. However, legitimate peripheral participation and more recent approaches place equal emphasis on the confidence aspect involved in critical thinking and complex learning (Kelller, 2010; Lave & Wenger, 1991; Milrad, Spector & Davidsen, 2003).

One approach, consistent with the foundations just reviewed, is to follow a path of graduated complexity and help the learner develop competence in a challenging problem domain (Milrad, Spector & Davidsen, 2003). Consistent with a graduated complexity approach is to use partially worked examples beginning with examples missing only a small part of an acceptable solution and then introduce examples with more and more parts missing eventually allowing a learner to develop an entire solution approach. In such cases, digital technologies can support such a process by knowing where a learner is within a learning progression path and introducing increasingly challenges examples to complete.

Yet another approach is to identify how experts typically think about that problem and use technology to encourage a learner to think more like an expert (Spector & Koszalka, 2004). This can be done by asking experts and learners to respond to four questions: (a) What key factors influence the problem situation? (b) How would you describe each of those factors? (c) What are the relationships among those factors? and (d) How would describe each of those relationships? Those responses can be put into the form of an annotated concept map and used as the basis for formative assessment.

These approaches are consistent with Dewey (1938), Lave and Wenger (1991), as well as with mainstream instructional design experts such as Gagné and Merrill (1990), Reigeluth (1999), and van Merriënboer (2012).
Gagné and Merrill (1990) proposed focusing on enterprises, which was called a whole task somewhat later by van Merriënboer (2012). Enterprises and whole tasks are authentic (representative of actual tasks) and consistent with the notion of problem-based learning (Barrows & Tamblyn, 1980). What is changing is the ability to provide real-time, meaningful feedback during problem solving using current and emerging technologies, which has been demonstrated by a few of the assessment technologies discussed in the next section.

**Current assessment technologies and systems**

Researchers have already accepted the importance of formative assessment in the teaching-learning process. In the 21st century, the availability and use of technological resources like computers, mobile devices, tablets, etc. are increasing rapidly in the school and higher education (Organisation for Economic Co-operation and Development (OECD), 2015). Now the question arises, “What kind of roles can technology play to support formative assessment?” For example, Bennett and Gitomer (2009) enumerated some of the important benefits of using technology, which are: (a) more informative, technology can help to track the full record of the problem-solving process adapted by the learner; (b) more efficient, saves time for scoring and error free; and (c) cost-effective, saves the expenses over the human scoring process. Spector et al. (2016) also advocated the application of current available technologies for formative assessment purposes. In this section, we discussed some of the currently available systems for formative assessment. In addition, we presented some empirical evidence supporting the key role played by different technologies in formative assessment we discussed some of the currently available systems for formative assessment.

While the gap addressed herein concerns the lack of formative assessment in complex problem-solving domains, there have been recent improvements in technologies, which can and have addressed that gap, at least in research settings. These technologies, however, have yet to see large-scale implementation in instructional settings, so the gap previously discussed remains. However, the means to systematically address that gap are now possible due to emerging technologies. Recently, many digital assessment tools have been developed, which may provide support to FA. These tools include HIMATT (Pirnay-Dummer, Ifenthaler, & Spector, 2010), AKOVIA (Ifenthaler, 2014), AssiStudy (Rodrigues & Oliveira, 2014), iSMILE (Bhagat, Subheesh, Bhattacharya, & Chang, 2017), etc.

HIMATT (Highly Integrated Model Assessment Technology and Tools) is a comprehensive tool which combines the features of DEEP (Dynamic Enhanced Evaluation of Problem Solving), MITOCAR (Model Inspection Trace of Concepts and Relations), T-MITOCAR (Text-MITOCAR), and SMD Technology (Structure, Matching, Deep Structure). HIMATT has two platforms: HIMATT Research Engine, which conducts and analyses experiments and HIMATT Subject Environment, which assigns the experiments to the individuals dynamically. The application of HIMATT includes states and changes, analysis and comparisons.

The framework of AKOVIA (Automated Knowledge Visualization and Assessment) is based on HIMATT. It is applicable for the semantic analysis of natural language (e.g., discussion forums, essay writing) and graphical knowledge representations. Automated feedback is one of the key features of AKOVIA, which can help the learners to understand their writing and improve it accordingly in an effective way.

AssiStudy is based on Service-Oriented Architectures (SOA). It creates personalized training exams based on students’ profile using question from the past exams stored in the repository. These training exams provide immediate feedback explaining the mistakes made by the students. This system used various Natural Language Processing (NLP) techniques to match reference answers (RA) with student answers (SA). After the training exams, teachers use students’ performance information to develop evaluation exams. Three main types of exams can be created: enumeration, specific knowledge and essay. Evaluation exams are checked by both AssiStudy and the teacher. This system tracks the performance of each student in formative as well as summative assessment and provides detailed pattern analysis to the teachers.

Identification of Students’ Misconceptions in Individualized Learning Environment (iSMILE) System is developed to provide feedbacks based on misconceptions in understanding a particular concept. This system is based on Model View Controller (MVC) architecture. Assessment procedure has two levels. Firstly, student needs to answer a root question. In the next step, a linked question is provided based on the answer for the root question to evaluate the deeper understanding of the concept. After finishing both levels, students are provided elaborated feedback about their misconceptions if they make any mistakes.

314
Hwang and Chang (2011) developed a Formative Assessment-Based Mobile Learning (FAML) system and evaluated on a local culture course learning. FAML used to provide only hints, when the students failed to find correct answers and motivated the learners to find the answers by their own. The results showed higher learning performance, learning attitude, and learning motivation by using FAML. In another study, an online game-based formative assessment named tic-tac-toe quiz for single-player (TRIS-Q-SP) was developed by Hooshyar et al. (2016). In addition, Hooshyar and colleagues (2016) integrated TRIS-Q-SP with an Intelligent Tutoring System (ITS) to teach computer programming. Three types of feedback (delayed feedback, knowledge result and elaborated feedback) were integrated within the system. The results revealed that TRIS-Q-SP improved experimental group’s problem-solving skills, which resulted into better learning achievement. More importantly, immediate elaborated feedback was one of the reasons for better learning performance. Conejo, García-Viñas, Gastón, and Barros (2016) employed a web-based assessment tool called Siette in a botany course for higher education. This system provided elaborated immediate feedback, which means the instant correct answer with detailed explanation for the wrong answers. They found that students who used Siette performed better than the students who underwent traditional method for formative assessment. More interestingly, immediate feedback helped the students for better performance, which is not possible without the use of technology. Recently, Faber, Luyten, and Visscher (2017) examined the effects of a digital formative assessment tool, Snappet on students’ mathematics achievement and motivation. This system also provides elaborated feedback about the errors made by the students and explanation for the correct answers. They also concluded that feedback contributed for the better performance and motivation of the students who used Snappet.

In addition, there is important research emerging with regard to the use of advanced technologies to support stealth assessment, which involves the use of log data in an online learning environment to determine areas in which a learner may be struggling and then help guide that learner to more productive outcomes (DiCerbo, Shute, & Kim, 2017; Shute & Moore, 2017). Moreover, the ability of digital technologies to collect data and present visualizations to help a learner make progress are also becoming more prominent in support of learning (Wang, Wu, Kinshuk, Chen, & Spector, 2013; Wu, Wang, Grotzer, Liu, & Johnson, 2016; Wu, Wang, Spector, & Yang, 2013; Yuan, Wang, Kushniruk, & Peng, 2016). Such visualizations often involve extensions of concept mapping techniques previously mentioned and address both content representations as well as feedback on student performance.

Looking to the future

Overall, research indicates that technology can support formative assessment to enhance learning performance, learning attitude, and learning motivation effectively across different disciplines. There is no doubt that technology can be used to support formative assessment, although technology has more often been used to provide access to and interaction with learning resources. Given the history of emphasis on formative assessment and the potential of new technologies to extend formative assessment into complex problem-solving domains, the potential for greater impact of formative assessment on the development of competence with regard to higher order learning is high. Promising technologies include stealth assessments, automated concept map based assessments, visualizations in support of formative assessment and self-regulation skills, and tools to promote networking and collaboration.

The conceptual framework we propose is to build on the following key notions:

- Continue to introduce real-world problems, however simplified, into curricula whenever possible;
- Build on the notions of graduated complexity (Milrad, Spector & Davidsen, 2003), enterprises (Gagné & Merrill, 1990), elaboration theory (Reigeluth, 1999), and whole tasks (van Merriënboer, 2012);
- Use annotated concept maps and causal influence diagrams as a means to elicit how someone is thinking about a complex problem;
- Compare progress towards expert-like thinking based on a series of problem conceptualizations.

Concluding remarks

In closing, it is our belief that formative assessment is an important albeit neglected task of educators. The primary job of a teacher is to help students succeed in their learning activities and educational pursuits. If one accepts that last remark, then one must continue to place emphasis on formative assessment and contribute to ongoing efforts to make effective use of new and emerging technologies in support of formative assessment, especially in complex problem-solving domains.
There are a number of ways to move forward with regard to a framework for supporting the development of complex problem-solving skills. These ways include the following:

- Develop repositories of representative complex problems in a variety of domains (this is being done with regard to a number of science learning tasks by the Smithsonian Institution – see https://ssec.si.edu/);
- Develop associated repositories of how experts think about those problems using annotated problem conceptualizations of the type found in DEEP and HIMATT (Pirnay-Dummer, Ifenthaler, & Spector, 2010; Spector & Koszalka, 2004);
- Develop a version of HIMATT that can be used in classroom and online course settings to provide dynamic, real-time feedback as a learner develops a conceptualization of the problem; and,
- Track and report the development of complex problem-solving skills because of using these technologies.

We have not reported new research in this short piece. Rather, we have been urging a particular perspective to support further research on the most promising technologies and learning approaches that have evolved in the last 50 years. What would be genuinely innovative and original would be to plan, implement, deploy on a large-scale version of the proposed approach and determine the impact on the development of critical thinking and complex problem-solving skills now being emphasized in so many places.

References


