Educational Technology & Society
An International Journal

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

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

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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

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


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

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

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

Associate editors
- Vladimir A Fomichev, K. E. Tsio1kovsky Russian State Tech Univ, Russia; Olga S Fomichova, Studio "Culture, Ecology, and Foreign Languages". Russia; Piet Kimmers, University of Twente, The Netherlands; Chul-Hwan Lee, Inchon National University of Education, Korea; Brent Muirhead, University of Phoenix Online, USA; Erkki Sutinen, University of Joensuu, Finland; Vladimir Uskov, Bradley University, USA.

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

Advisory board
- Ignacio Aedo, Universidad Carlos 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; Tae-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Harley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jemni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Rob Koper, Open University of the Netherlands, The Netherlands; Jimmy Ho Man Lee, The Chinese University of Hong Kong, Hong Kong; Ruddy Lelouche, Université Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Roy McGreal, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Växjö University, Sweden; Riichiro Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoshi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shih, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J. Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Chin-Chung Tsai, National Taiwan University of Science and Technology, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J.H. Yang, National Central University, Taiwan; Yu-Mei Wang, University of Alabama at Birmingham, USA.

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

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

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

Abstracting and Indexing

Guidelines for authors
Submissions are invited in the following categories:
- Peer reviewed publications: Full length articles (4000 - 7000 words)
- Book reviews
- Software reviews
- Website reviews

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

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

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

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

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

Please provide following details with each submission: • Author(s) full name(s) including title(s), • Name of corresponding author, • Job title(s), • Organisation(s), • Full contact details of ALL authors including email address, postal address, telephone and fax numbers.

The submissions should be uploaded at http://www.ifets.info/ets_journal/upload.php. In case of difficulties, please contact kinshuk@ieee.org (Subject: Submission for Educational Technology & Society journal).
# Journal of Educational Technology & Society

**Volume 18 Number 2 2015**

## Table of contents

### Special Issue Articles

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guest Editorial: Technology Supported Assessment in Formal and Informal Learning</td>
<td>1–2</td>
</tr>
<tr>
<td>Jesús García Laborda, Demetrios G Sampson, Ronald K. Hambleton and Eduardo Gazman</td>
<td></td>
</tr>
<tr>
<td>Tablet-Based Math Assessment: What Can We Learn from Math Apps?</td>
<td>3–20</td>
</tr>
<tr>
<td>Gabrielle A. Cayton-Hodges, Gary Feng and Xingyu Pan</td>
<td></td>
</tr>
<tr>
<td>Using Triologues to Measure English Language Skills</td>
<td>21–32</td>
</tr>
<tr>
<td>Youngsoon So, Diego Zapata-Rivera, Yeonsuk Cho, Christine Luce and Laura Battistini</td>
<td></td>
</tr>
<tr>
<td>M-AssIST: Interaction and Scaffolding Matters in Authentic Assessment</td>
<td>33–45</td>
</tr>
<tr>
<td>Patricia Santos, John Cook and Davinia Hernández-Leo</td>
<td></td>
</tr>
<tr>
<td>A Web-Based Course Assessment Tool with Direct Mapping to Student Outcomes</td>
<td>46–59</td>
</tr>
<tr>
<td>Walid Ibrahim, Yacine Atif, Khaled Shuaib and Demetrios Sampson</td>
<td></td>
</tr>
<tr>
<td>A Reconfigurable Simulation-Based Test System for Automatically Assessing Software Operating Skills</td>
<td>60–79</td>
</tr>
<tr>
<td>Jun-Ming Su and Huan-Yu Lin</td>
<td></td>
</tr>
<tr>
<td>A Study of Perceptual Typologies on Computer Based Assessment (CBA): Instructor and Student Perspectives</td>
<td>80–96</td>
</tr>
<tr>
<td>Jin-Young Kim</td>
<td></td>
</tr>
<tr>
<td>Antonio Robles-Gómez, Salvador Ros, Roberto Hernández, Llanos Tobarra, Agustín C. Caminero and José M. Agudo</td>
<td></td>
</tr>
<tr>
<td>Group Learning Assessment: Developing a Theory-Informed Analytics</td>
<td>110–128</td>
</tr>
<tr>
<td>Wanli Xing, Robert Wadholm, Eva Petakovic and Sean Goggins</td>
<td></td>
</tr>
</tbody>
</table>

### Full Length Articles

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does Digital Scholarship through Online Lectures Affect Student Learning?</td>
<td>129–139</td>
</tr>
<tr>
<td>Shelley Kinash, Diana Knight and Matthew McLean</td>
<td></td>
</tr>
<tr>
<td>A Two-Stage Multi-Agent Based Assessment Approach to Enhance Students’ Learning Motivation through Negotiated Skills Assessment</td>
<td>140–152</td>
</tr>
<tr>
<td>Abdelhafid Chadli, Fatima Bendella and Erwan Tranvouez</td>
<td></td>
</tr>
<tr>
<td>An Exploration into Improving Examinees’ Acceptance of Participation in an Online Exam</td>
<td>153–165</td>
</tr>
<tr>
<td>I-Fan Liu, Ruey-Shin Chen and Hao-Chun Lu</td>
<td></td>
</tr>
<tr>
<td>Apply an Augmented Reality in a Mobile Guidance to Increase Sense of Place for Heritage Places</td>
<td>166–178</td>
</tr>
<tr>
<td>Yu-Lien Chang, Huei-Tse Hou, Chao-Yang Pan, Yao-Ting Sung and Kuo-En Chang</td>
<td></td>
</tr>
<tr>
<td>Conceptual Tutoring Software for Promoting Deep Learning: A Case Study</td>
<td>179–194</td>
</tr>
<tr>
<td>Angela Stott and Annemarie Hattingh</td>
<td></td>
</tr>
<tr>
<td>Students’ Groupwork Management in Online Collaborative Learning Environments</td>
<td>195–205</td>
</tr>
<tr>
<td>Jianzhong Xu, Jianxia Du and Xitao Fan</td>
<td></td>
</tr>
<tr>
<td>From Learning Object to Learning Cell: A Resource Organization Model for Ubiquitous Learning</td>
<td>206–224</td>
</tr>
<tr>
<td>Shengquan Yu, Xianmin Yang, Gang Cheng and Minjuan Wang</td>
<td></td>
</tr>
<tr>
<td>Analyzing Log Files to Predict Students’ Problem Solving Performance in a Computer-Based Physics Tutor</td>
<td>225–236</td>
</tr>
<tr>
<td>Young-Jin Lee</td>
<td></td>
</tr>
</tbody>
</table>
The Comparison of Solitary and Collaborative Modes of Game-based Learning on Students’ Science Learning and Motivation

Ching-Huei Chen, Kuan-Chieh Wang and Yu-Hsuan Lin

The Role of ICT Infrastructure in Its Application to Classrooms: A Large Scale Survey for Middle and Primary

Chun Lu, Chin-Chang Tsai and Di Wu

Impeding Phenomena Emerging from Students’ Constructivist Online Game-Based Learning Process: Implications for the Importance of Teacher Facilitation

Morris Siu-yung Jong and Junjie Shang

A Learning Style-based Grouping Collaborative Learning Approach to Improve EFL Students’ Performance in English Courses

Yu-Chen Kuo, Hai-Chun Chu and Chi-Hao Huang

Learning Mathematics with Interactive Whiteboards and Computer-Based Graphing Utility

Ayhan Kursat Erbas, Muge Ince and Sukru Kaya

Effects of Verbal Components in 3D Talking-head on Pronunciation Learning among Non-native Speakers

Ahmad Zamzuri Mohamad Ali, Kogilathah Segaran and Tan Wee Hoe

Applying The CHAID Algorithm to Analyze How Achievement is Influenced by University Students’ Demographics, Study Habits, and Technology Familiarity

Bahar Baran and Eylem Kilic

Exploring Elementary-School Students’ Engagement Patterns in a Game-Based Learning Environment

Ya-Hui Hsieh, Yi-Chun Lin and Huei-Tse Hou

Understanding the Self-Directed Online Learning Preferences, Goals, Achievements, and Challenges of MIT OpenCourseWare Subscribers

Curtis J. Bonk, Mimi Miyoung Lee, Xiaojing Kou, Shuya Xu and Feng-Ru Sheu

The Multimedia-Based Learning System Improved Cognitive Skills and Motivation of Disabled Children with a Very High Rate

Sawsan Saad, Amal Dandashi, Jihad M. Aljaam and Moataz Saleh

Revisiting the Blended Learning Literature: Using a Complex Adaptive Systems Framework

Yuping Wang, Xibin Han and Juan Yang

Book Reviews

Book review: Active Learning Spaces: New Directions for Teaching and Learning (Author: Paul Baepler et al.)

Mingze Sun and Feng-Kuang Chiung
Guest Editorial: Technology Supported Assessment in Formal and Informal Learning

Jesús García Laborda, Demetrios G Sampson, Ronald K. Hambleton and Eduardo Guzman

University of Alcalá, Alcalá de Henares-Madrid, Spain // University of Piraeus & Center for Research and Technology, Hellas, Greece // Educational Assessment, University of Massachusetts, Amherst, USA // University of Malaga, Malaga, Spain // jesus.garcialaborda@uah.es // sampson@iti.gr // rkh@educ.umass.edu // guzman@lcc.uma.es

Assessment can be considered as a systematic process of making judgments, and consequently reporting results, either about the effectiveness (with respect to achievement of intended learning outcomes) of learning and educational processes (ranging from formally established educational programs to informal learning experiences), or about individual students’ progress toward attainment of established educational objectives. Technology holds the potential to facilitate assessment implementation and maximize the benefits for all involved stakeholders. Cases of technology-enabled assessment are traced back to the use of the abacus and writing techniques for evaluating knowledge acquisition more than 2000 years ago. However, the use of technology alone, without being followed by the necessary paradigm shifts, cannot result to a significant effect in learning and teaching. The sometimes divergent evidence reported indicate the need for re-conceptualizing the ways in which technology should support both learning and assessment. Today, typically, learners and teachers use technology for accessing learning resources or submitting assigned homework, whereas technology has the potential to facilitate engagement in meaningful and authentic learning experiences and methods of keeping track of their levels of achievement. Technology is a vehicle that has the potential to help toward effectively meeting learning and assessment needs. However, this potential cannot be realized without taking account of the fact that learning and consequently assessment takes place within the context (both social and technological) in which the learner acts. Appropriately designed methods and tools may facilitate monitoring of learning processes and, with the necessary technology-led and/or instructor-led feedback and scaffolds, learners are able to reflect and adjust their actions toward optimizing the outcomes of their efforts by performing even beyond their Zone of Proximal Development.

A widely adopted field of application of technology for assessment purposes is that of computer-based testing, with appropriately designed software tools being used for generating and administrating tests, as well as reporting results. Ease of test items development, reduced costs of administration to large numbers of examinees, increased levels of tests validity and reliability, and the capacity to communicate results in a variety of formats to a range of stakeholders constitute a number of reported advantages. Based on Item Response Theory, there is potential to generate tests of different difficulty levels, as well as adapt the difficulty level of assigned test items at the run time level depending on performance. Technology allows flexibility in time scheduling with learners being able to take tests anywhere and in any time. Thus, the act of assessment may not be constrained either by the physical boundaries of the traditional classroom or by the stiff school/university timetable. Additionally, there is potential for accommodating people with physical disabilities by providing appropriate input and output means.

However, testing constitutes only one of the available category of assessment methods, and its use for grading and learners’ ranking only one of the many purposes of assessment. Despite the bespoken advantages of computer-based testing, technology can reveal its real potential when used for the execution of assessment activities that are seamlessly interwoven in either formal or informal learning processes; that is, when used for formative assessment rather than summative assessment purposes. By this way both learners and instructors may have access to performance data in-context that will help eliciting decisions about future learning directions. Simulations, digital games, virtual worlds, virtual labs, as well as the use of e-portfolios, constitute characteristic examples of technological facilitators of formative, in-situ, assessment. From this perspective, current trends in computer-based testing, technology-supported assessment in MOOCs, competence-based assessment methods and tools, and innovative approaches regarding generation and administration of new types of test items and constructs, as well as validation processes, need to be considered within the context of technology-supported assessment (Ifenthaler, 2014; Fisteus, Pardo, & García, 2013; Miller, Baker & Rossi, 2014; Webb, Gibson & Forkosh-Baruch 2013).

To this end, this special issue targets at efforts that offer insights into current and future trends and research directions in technology-based assessment and testing. It attracted 32 submissions which where double blindly...
reviews by 42 international experts. Finally, 8 papers have been selected for publication covering a wide range of topics in this field.

Gabrielle A. Cayton-Hodges, Gary Feng and Xingyu Pan, in their contribution entitled “Tablet-based Math Assessment: What Can We Learn From Math Apps?” present a survey of mathematics education apps in the Apple App Store, conducted as part of a research project to develop a tablet-based assessment prototype for elementary mathematics.

Youngsoon So, Diego Zapata-Rivera, Yeonsuk Cho, Christine Luce and Laura Battistini, in their contribution entitled “Using Trialogues to Measure English Language Skills” explore the use of technology-assisted, trialogue-based tasks to measure the English language proficiency of students learning English as a second or foreign language.

Patricia Santos, John Cook and Davinia Hernández-Leo, in their contribution entitled “m-AssIST: Interaction and Scaffolding matters in authentic assessment” presents the m-AssIST model which captures the essential emerging properties to design and analyse m-assessment activities. Their model is used to analyse the benefits and limitations of existing m-test based systems, and the authors discuss the importance of meaningful interactions, and the provision of scaffolding mechanisms to support formative and authentic assessment supported by mobile technologies.

Walid Ibrahim, Yacine Atif, Khaled Shuaib and Demetrios Sampson, in their contribution entitled “A Web-Based Course Assessment Tool with Direct Mapping to Student Outcomes” a novel course assessment process supported by a Web based interface that articulates and streamlines the assessment data collection, performance evaluation and tracking of remedial recommendations in the context of higher education.

Jun-Ming Su and Huan-Yu Lin, in their contribution entitled “A Reconfigurable Simulation-Based Test System for Automatically Assessing Software Operating Skills” present a reconfigurable simulation-based test system designed on a Software Operation-Finite State Machine targeting to reduce cost and time effort through sharing and reusing the designed simulation-based tests items.

Jin-Young Kim, in his contribution entitled “A Study of Perceptional Typologies on Computer Based Assessment (CBA): Instructor and Student Perspectives” reports on different viewpoints on Computer Based Assessment (CBA) by using Q methodology to identify perspectives of students and instructors and classify these into perceptional typologies.

Antonio Robles-Gómez, Salvador Ros, Roberto Hernández, Llanos Tobarra and Agustín C. Caminero, in their contribution entitled “User Acceptance of a Proposed Self-Evaluation and Continuous Assessment System” present a service-oriented system that follows students’ progress and automatically assess their practical abilities, through self-evaluation and continuous assessment.

Wanli Xing, Robert Wadholm, Eva Petakovic and Sean Goggins, in their contribution entitled “Group Learning Assessment: Developing a Theory-Informed Analytics,” present a study that explores natural language processing of the chat log data and incorporates it into an activity theory measure construction system and visualizations in order to further inform learning assessment in Computer Supported Collaborative Learning (CSCL).

References


Table-Based Math Assessment: What Can We Learn from Math Apps?

Gabrielle A. Cayton-Hodges1*, Gary Feng1 and Xingyu Pan2

1Educational Testing Service, Princeton, NJ 08541, USA // 2Cognitive Science Research Group, University of Michigan, Ann Arbor, MI 48109, USA // gcayton-hodges@ets.org // gfeng@ets.org // xypan@umich.edu

*Corresponding author

ABSTRACT

In this report, we describe a survey of mathematics education apps in the Apple App Store, conducted as part of a research project to develop a tablet-based assessment prototype for elementary mathematics. This survey was performed with the goal of understanding the design principles and techniques used in mathematics apps designed for tablets. We focused our reviews on four areas, (1) the quality of mathematical content, (2) feedback and scaffolding, (3) richness of interactions, and (4) adaptability of the applications. These four areas were developed from prior research on digital tools in mathematics (e.g., Digital Tools for Algebra Education criteria; Bokhove & Drijvers, 2011), designing principles of learning objects (e.g., Learning Object Evaluation Metric; Kay & Knaack, 2008), as well as quality of mathematics instruction (e.g., Hill et al., 2008). We end with recommendations for tablet assessment design cultivated through this review.

Keywords

Technology, Tablets, Formative assessment, Elementary school, Mathematics

Introduction

The era of tablet computers as wide-spread consumer devices began in 2010 with the announcement of the Apple iPad. In 3 years, the tablet shipment has surpassed either desktop or laptop PCs, and outsold all PCs combined in the fourth quarter of 2013 (IDC, 2013). Tablets see even stronger demands in the education sector (Interactive Education Systems Design, Inc., 2013). In the educational assessment arena, tablet devices have been recognized as an alternative test delivery platform, and various validity issues have been studied (see, for example, Laughlin Davis, Strain-Seymour, & Gay, 2013; Strain-Seymour, Craft, Davis, & Elbom, 2013). How to design educational assessments specifically for tablet computers, however, remains largely an uncharted territory.

Change is imminent, though. Decisions by prominent testing consortia such as PARCC and Smarter Balanced to support touch-screen devices have pushed tablet-based assessments into the spotlight (see Smarter Balanced Assessment Consortium, 2012). High stakes tests are not the only impetus for educators, students, and parents when they order tablets by the thousands (see Blume, 2013); instead, they are drawn to the numerous possibilities that tablets hold, including the idea that tablets will personalize learning and improve student learning outcomes. A key to achieving this goal is in the form of formative assessments to track student learning and provide timely feedback. In sum, tablet-based assessments – both summative and formative – are scheduled to launch shortly. Are we ready?

The answer depends on one’s expectations. If the question is whether we are ready for tablet delivery of assessments that were designed for paper and pencil or PCs, the answers, and debates, can be found in various technical analyses and usability studies (e.g., Isabwe, 2012; Kowalski, Kowalski, & Gardner, 2009; Kim, Lim, Choi, & Hahn, 2012; Toby, Ma, Lai, Lin, & Jaciw, 2012; Laughlin Davis et al., 2013; Tsuei, Cho, & Chan., 2013; Strain-Seymour et al., 2013). The consensus seems to be that the technology can be made to work. But if the goal is for tablet-based assessments to be as engaging and interactive as the many apps that students are already enjoying on tablets, we have a long way to go.

This report is a part of a research project under the Cognitively Based Assessment of, for, and as Learning (CBAL) initiative at ETS that aims to build a model for an innovative K–12 assessment system using learning sciences and other related research towards: documenting what students have achieved (of learning); helping to identify how to plan instruction (for learning); and being a worthwhile educational experience in and of itself (as learning, see Bennett, 2010). This survey was inspired by work of the first author to develop a tablet-based assessment prototype for elementary mathematics, namely, students’ understanding of fractions and decimals. When we began the project, we found that most published research on the use of tablets in mathematics assessments focused either narrowly on specific features of the tablet technology, such as the usability of digital ink (e.g., Ren & Moriya, 2000; McKnight & Fitton, 2010; Kim et al., 2012), or broadly on the relationship between classroom tablet usage and end-of-semester student learning outcomes (Hieb & Ralston, 2010; Isabwe, 2012; Kowalski et al., 2009). Given the apparent lack of a
systematic review on how to design mathematics assessments for tablet computers, we decided to start with a survey of mathematics education apps with the hope of understanding their design principles and techniques. We also looked for potential pitfalls that we hope to avoid in our work. This report is a summary of this qualitative research.

In what follows we will outline the dimensions of mathematical content and interactions used as the basis for review. We then outline the method we used to sample the apps and the evaluation criteria. In the Results section we summarize the main findings from our survey. The Discussion section focuses on the lessons learned from the survey and key take-aways for design of mathematics assessment and instruction applications (i.e., what works and how we can apply these techniques in our tablet-based assessment design).

**Dimensions of mathematical review**

We focused our review on four dimensions: (1) the quality of mathematical content, (2) feedback and scaffolding, (3) richness of interactions, and (4) adaptability of the applications. These four areas were cultivated from prior research on digital tools in mathematics (e.g., Digital Tools for Algebra Education criteria; Bokhove & Drijvers, 2011), from the evaluation of mobile apps in algebra (Harrison, 2013), and from designing principles of learning objects (e.g., Learning Object Evaluation Metric; Kay & Knaack, 2008), as well as from the quality of mathematics instruction (e.g., Hill et al., 2008). These dimensions are described below.

**Mathematical content**

We looked specifically at two main dimensions of mathematical content, derived from the Mathematical Quality of Instruction (MQI; see Hill et al., 2008): mathematical accuracy and mathematical richness. Mathematical accuracy is adapted from the MQI dimensions of “error” and “imprecision.” Moyer-Packenham, Salkind, and Bolyard (2008) define mathematical accuracy as “the degree to which the mathematical object is faithful to the underlying mathematical properties of that object in the virtual environment” (p. 204). In this case, we’re looking at whether the mathematical content in the app is true to its pure mathematical form, and, if not, where does the inaccuracy lie. The richness of mathematics includes two elements: attention to the meaning of mathematical facts and procedures and engagement with mathematical practices (Hill et al., 2008). A mathematically rich experience captures the meaning of the mathematical practices and provides links, explanations, and generalizations.

**Feedback and scaffolding**

An essential condition for supporting student learning and improving chances of success is to provide feedback and scaffolding (Gibbs & Simpson, 2004; Hattie & Timperley 2007; Nicol & MacFarlane-Dick 2006). We specifically looked at three dimensions: feedback, scaffolding, and opportunity for reflections. The Feedback dimension, derived from Digital Tools for Algebra Education criteria (Bochove & Drijvers, 2010), includes whether any feedback is given, whether the feedback is relevant to the input and the content, the timeliness of the feedback, and the types of feedback (e.g., conceptual, procedural, corrective).

The Scaffolding and the Opportunity for Reflection dimensions were derived from the Learning Object Evaluation Instruction (Haughey & Muirhead, 2005), illustrating “opportunities to extend and expand learning activities to beyond the confines of the object itself”. The Scaffolding dimension includes whether the app provides any scaffold to help student learning and, if so, the forms of scaffolds (e.g., hints and guiding questions). The Opportunity for Reflection dimension focuses on whether students have opportunities to think about or explain their thought process. This reflection may occur in the form of verbal or typed explanation, transfer to other problems, or answering questions related to the solution process.

**Richness of interactions**

Meaningful interactions between students and the apps not only contribute to student learning outcomes (Chan & Black, 2006), but also influence the quality as well as the interpretation of the students’ actions. We focused on the
two elements of interactions, derived from the assessment criteria in Digital Tools for Algebra Education (Bokhove & Drijvers, 2010): the modes of interactions and the item types offered.

Modes of interactions may include training, practice test, actual test, and others (Bochove & Drijvers, 2010). Modes are used to organize selective presentation of different attributes within the same app. For example, scaffolding and hints may be available only in training and practice test modes but not in the actual test mode. An app may allow different item types, from static multiple-choice questions to dynamic simulation of real world settings where students interact realistically with data of considerable complexity within a multimedia environment. Different item types allow different opportunities to observe and assess student behavior and reasoning (Gorin, 2006). In this case, we're looking at all interactions present during app use including interactions with the mathematics, the scenario, the digital features, as well as the device.

**Scoring and adaptability**

Scoring of students’ performance is a central component of assessment development. Scores may be used to evaluate student performance and provide real-time feedback for the student, yet the meaning of the scores depends on how the scores are calculated. For instance, the meaning of a score is different when only accuracy is considered as compared with when both accuracy and speed are considered. Scores may also be used as part of a student profile and mastery account to administer appropriate questions (adaptability), while the adaptability will also influence the scoring algorithm. We focus on two elements within this category, namely scoring method and adaptability. Scoring method focuses on which variables were included in score calculation within each app (e.g., correct/incorrect, number of hints needed, time, etc.), while adaptability concerns whether and how each app provides user-dependent content (e.g., on-demand hints vs. on-error hints, adapting to user ability vs. always increasing difficulty, etc.).

**Method**

**Sample**

We decided to focus on mathematics education apps on the Apple App Store for two reasons. First, the Apple iOS platform is the most popular tablet platform for the target population (4th- to 5th-grade students, see Mainelli, 2013) and second, the Apple App Store contains many more apps as well as user reviews than other platforms.

We sampled the mathematics education apps on the Apple App Store during the summer of 2013 using the following approach. First, we obtained a list of the top 100 most popular education apps, of which 12 were mathematics oriented. Then we obtained the names of an additional 52 apps that were featured in either the “education collection” for math in iTunes or iTunes Education Spotlight (http://www.apple.com/itunesnews/education/us/).

From this list of 64 apps, decisions to review an app were based on a number of factors: calculator apps were excluded, apps with few downloads and/or low ratings from users were excluded, and apps that were obvious clones of other apps were excluded. Additionally, some apps were offered in series, with each individual app featuring one topic (e.g., algebra, fraction, counting, etc.). For each series, only the most popular app (i.e., the one with the most downloads) was included. Finally, we excluded some paid apps with a high cost for practical reasons of budget limitations. This process resulted in a sample of 16 apps downloaded for review (see Appendix for the list of apps reviewed).

**The review process**

Selected apps were downloaded and installed on an iPad device and a single researcher reviewed all sixteen apps, allowing 10 - 25 minutes of interaction with each app, depending on the complexity of the design and the variability of interactions available. Additional information regarding each app was collected from user ratings and written comments on the App store. Video reviews of each app were also collected through a keyword search on Youtube.com. A second researcher then reviewed the sixteen apps, along with all notes for general agreement on comments and interactions.
Notes were taken according to the four dimensions outlined in the Introduction. Notes included the positive and negative aspects of the app with regards to each dimension. Because our primary goal is to learn from tablet-based mathematics apps in general rather than to evaluate this particular sample of apps, we did not pursue quantitative analyses of the notes. We will present a qualitative summary of the lessons learned from our interactions with the apps.

Results

Range of mathematics apps

The initial sample of sixty-four applications, as well as the sixteen-app subset reviewed, covered a wide range of mathematics topics, including numbers and operations, algebra, geometrics, and statistics and probability. The target age of the applications ranges from preschool to adults, with most apps focusing on younger populations, namely from preschool to elementary school-aged children.

The apps also differ in genres. Some apps take the form of e-textbooks, providing instructional materials as well as unit-based assessments (e.g., HMH Fuse series, Woot Math). Others take the form of a personal tutor, offering video clips or demonstrations of procedures (e.g., Long Division; Khan Academy). The vast majority of apps, however, take the form of games, in which the player needs to solve mathematics problems to earn points and achieve game goals (e.g., DragonBox Algebra, Motion Math series, Teachley; Addimal Adventure, etc.).

Interestingly, we did not find any apps that claim to be assessment apps, notwithstanding some that are test-prep in nature. On the other hand, both the e-textbook and the game applications typically have built-in assessment elements. Therefore the current review will still provide insights for building tablet-based math assessments.

Evaluation of selected mathematics apps

Mathematical content

When reviewing the content of the sixteen selected apps, we looked specifically at two main dimensions, derived from the Mathematical Quality of Instruction (MQI; see Hill et al., 2008): mathematical accuracy and mathematical richness.

Mathematical accuracy

As stated earlier, mathematical accuracy is adapted from the MQI dimensions of “error” and “imprecision.” Moyer-Packenham et al., (2008) define mathematical accuracy as “the degree to which the mathematical object is faithful to the underlying mathematical properties of that object in the virtual environment” (p. 204).

The mathematics apps appear mostly accurate in mathematical content, but sometimes conscious design decisions are made to sacrifice mathematical accuracy for ease of use or to match the user expectations. For example, DragonBox Algebra, is designed to teach students how to solve for the value of an unknown. Certain valid mathematical solutions are not possible, however, depending on the level of the game, presumably to eliminate distractions for students performing at lower levels. For example, sometimes both sides of the equation can be multiplied by the same number, but not divided. Another type of example can be seen in the problem: $x \times x = x/3$. The user is expected to divide both sides of the equation by $x$, obtaining the solution $x = 1/3$. This, however, makes it impossible to get the alternative [valid] solution of $x = 0$. These and other examples may generate or reinforce mathematical misunderstandings.

Mathematical richness

The richness of mathematics includes two elements: attention to the meaning of mathematical facts and procedures and engagement with mathematical practices (Hill et al., 2008).
Meaning making includes explanations of mathematical ideas and drawing connections among different mathematical representations. Most applications are doing poorly in this aspect. They typically focus on the retrieval of mathematical facts using simple response item types (such as numeric response consisting of a single digit) with automatically generated items. They may also focus on how to carry out procedures (e.g., fraction addition with different denominators, long division, etc.). We did not find any apps in our review that engaged users in activities where they need to explain or reflect on why a procedure works or why certain strategies do not work.

As for drawing connections among different representations, apps vary greatly. Some apps offer only Arabic numerals and algebraic expressions (e.g., TouchyMath by Joel Martinez), while others present multiple representations without explicitly drawing connections between them. For example, DragonBox Algebra offers graphic demonstration of the additive inverse relationship. An object icon and its color-inverted counterpart will cancel each other once they are put on top of each other. It also offers numeric expressions of additive inverse relationships, such as $5 + (-5) = 0$. No explicit connections, however, are drawn between the graphic and the numerical representation. This is part of the design philosophy of DragonBox Algebra, which claims to teach rules of algebra surreptitiously in a game without explicit references to numbers and variables. In contrast to the above, some apps focus explicitly on mappings among different mathematical representations. For example, in Motion Math HD: fraction, users need to identify identical fractions represented in different forms, including area models, number line models, fractions, percentages, and decimals.

Overall, most apps do poorly on engaging users in meaningful mathematical practices. Mathematical practices include the presence of multiple solution methods, the development of mathematical generalizations from specific examples, and the fluent and precise use of mathematical language (Hill et al., 2008). In our search, we were unable to find any apps on the Apple App store that require users to apply multiple strategies to solve a problem. Efficiency of the solution is usually evaluated through the speed of users’ responses. But some apps adopt a different approach and evaluate the efficiency of the solution by the number of moves used to solve a problem. For example, DragonBox Algebra has set a maximum number of moves for each item. The user will have to isolate the unknown variable on one side of the equation before he/she runs out of moves. In this way, users need to solve each problem in a relatively efficient way, rather than through a trial-and-error approach.

In sum, our impression is that most apps we reviewed provide relatively accurate mathematical content, yet, in one way or another, most apps are inadequate in the mathematical richness dimension. This may be due in part to limitations of the genre of the mathematics apps – in a game that emphasizes speed, there is often little room for multiple solutions or reflections. In other cases the narrow focus may be a design decision made implicitly (e.g., the author may be unaware of the richness) or explicitly (e.g., DragonBox Algebra, which insists on not making explicit connections between in-game actions and algebraic operations).

Feedback and scaffolding

Although opportunities for reflection are few and far between in current mathematics education apps, feedback and scaffolding are available on most of them, taking on different forms.

Relevant and timely feedback

Feedback on user performance can take three forms: (a) It can give status feedback on the problem being solved or on the mathematical objects being examined; (b) it can be corrective feedback, letting the user know a mistake has been made and guiding the user to correct the mistake; (c) it can be conceptual feedback, asking the user questions that cause the user to reconsider his or her perceptions of the objects.

A majority of the apps explored offer corrective feedback upon finishing an item. For example, Algebra Touch by Regular Berry Software LLC clearly indicates when an expression is fully simplified or an equation is solved. Invalid actions are immediately identified visually, audibly, and tactiley as parts of the equation vibrate to indicate where and why the action was invalid. As for status feedback, a score board/bar is present in most apps to provide information about a user’s overall performance so far. Users may also choose to go to a status board to review their performance on each item.
Compared with the corrective and status feedback, conceptual feedback is less common. For example, in *DragonBox Algebra*, invalid actions are visibly and audibly identified (although as discussed before, some otherwise valid actions are not allowed), but reasons for their invalidity are not given. On the other hand, some apps provide conceptual feedback during the user’s problem-solving experience. For example, *BuzzMath Middle School* will offer students a complete problem-solving procedure and explanation when they answer questions incorrectly. Some apps will not only provide feedback on what is the right solution and why it is right, but also provide feedback on why the given solution is wrong. In *TouchyMath* by Joel Martinez, when invalid actions are attempted, the app highlights terms or operations to indicate why the actions were invalid. When a user submits a solution that is not yet simplified, the app will indicate what parts of the solution can still be simplified. Conceptual feedback can also take the form of video lectures. For example, *Khan Academy*, which works as a video library and an extension of its website, syncs user activity across the platform, and links the user to video lectures of topics on which there was poor performance.

**Scaffolding**

Some applications offer scaffolding when students have consistently experienced difficulties. The scaffolding is typically in the form of hints or guiding questions that are offered when students demonstrate evidence of difficulty in problem solving, such as prolonged reaction time, or an incorrect first attempt. Hints are also offered when students request them (on-demand hints), regardless of their current performance. An example of an on-demand hint can be seen in *Pizza Fractions* by Brian West, in which students can choose to “peek at pizzas” to see the area model of the magnitude of fractions before they indicate their choice (Figure 1a). On-demand hints may not be tailored towards the specific item. For example, in *Numerosity: Play with multiplications*, a user’s request for a hint will activate a generic multiplication table. An example of as-needed hints comes from *Teachley: Addimal Adventure* by Teachley (Figure 1b). As shown in the picture, the block 2+5 is hanging by two chains. One chain is cut if no response is given after 2 seconds. Once one chain is cut, the hint for that problem jumps out. The hint on this picture suggests that the user should apply the “count on” strategy; for users unfamiliar with this strategy, a tutorial is available outside of the gaming interface.

![Figure 1. Example of hints in Pizza Fractions (a) and Teachley: Addimal Adventure (b)](image)

**Opportunities for reflection**

The apps we reviewed rarely promoted reflections through sense-making questions to stimulate user’s summarizing of mathematical rules. Some apps, however, do make available a complete history of users’ problem-solving actions at the end of the task (e.g., *TouchyMath*). The purpose of this is to afford students an opportunity to review their
problem-solving process. Without an explicit benefit or gain in the context of the game though, users are likely to skip this step and proceed to the next task.

**Richness of interactions**

The number of interactions available sets the limit for the possible user actions while using an app. Available interactions also influence how students approach a task. For example, different modes (i.e., studying, practice, or test) would influence students’ strategy choice as well as their level of engagement (Roediger, Agarwal, McDaniel, & McDermott, 2011; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011). Moreover, item types (e.g., multiple choice, short answers, essays) also influence how students approach tasks (Butler and Roediger, 2007).

**Modes of interactions**

By modes of interactions we mean phases of a game or an application that elicit distinct patterns of user interactions, for examples, tutorials, practices, competitions or main task sets, reviews or scoreboards, etc. Each mode has a different learning goal.

![Image](image.png)

*Figure 2. Example set up of practice questions in DragonBox Algebra*

Some apps offer practice questions via a game tutorial for students to get familiar with the types of interactions that will be used in the game. For example, *DragonBox Algebra* provides a tutorial to introduce the set-up of the screen, the different elements of the game (e.g., the “zero” card, the “one” card, the “dragon” card), as well as the goal of the game (i.e., isolate the “dragon” box on one side; see Figure 2). Users need to follow the instruction to finish a task before they start freely interacting with the game.

The practice mode is also used before or in mid-game, typically when students have difficulty with certain problems. For example, in *Hands-on Equation*, students may choose to watch a short video-clip for help. After watching the video, they need to finish two practice problems before getting back to the “exercise” interface.
Item types

In most of the apps there are fairly clear units of interactions that are analogous to test items in an assessment. They may be in the form of arithmetic problems to be solved one after another, or challenges to be passed in order to advance to the next level.

A great variety of items are available in these apps, although constructed response items are uncommon. Most items are variations of multiple-choice questions. Some items vary the content of the choices; instead of numbers, equations, or texts, choices may include pictures, graphs, or even video clips. Variations are also included in the layout of the choices. Instead of having a finite number of choices horizontally or vertically aligned, the number of choices may be infinite, (e.g., locations on a number line). Moreover, variations may also happen with the choice-making actions. Some choice-making movement involves novel interactions other than tapping. For example, in the Motion Math series, students need to tilt the tablet for the object to fall on the choice area located on the two lower corners of the screen. Drag-and-drop actions are also used in multiple-choice questions, where users need to drag the selected choices to desired locations.

Some apps use a combination of grid-in and multiple-choice format in presenting their questions. For example, in Numerosity: Play with Multiplication!, users may encounter vertical multiplication of multi-digit numbers. They need to select the number for the units first. The choices will refresh after that and they can select the number for tens (see Figure 3), then hundreds and thousands, if applicable.

Figure 3. Example of combining grid in with multiple-choice in Numerosity: Play with Multiplication!

Constructed response items usually involve interactions with virtual manipulatives, such as fraction strips and place value blocks. Items may require students to use manipulatives to represent mathematical entities. For example, in Hands on Math: Base Ten Blocks, users can drag unit blocks, ten blocks, and hundred blocks into virtual “trays” to represent multi-digit numbers. Items may also require students to use manipulatives to represent mathematical operations. For example, Figure 4 comes from Woot Math by Nimbee LLC. To solve this problem, users need to drag corresponding unit fraction “slices” into the “fraction circle”, in this case 3 “1/10” slices and 2 “1/5” slices, to show the process of 3/10 + 2/5. Moreover, users will also need to drag 2 “1/10” slices on top of each “1/5” slice to simulate the process of finding equivalent fractions.

Handwritten input was rare among the reviewed apps. When it was used, users’ handwritten input may be left “as is” (e.g., Woot Math) or translated into text input for further processing (e.g., FluidMath, Todo K-2 Math Practice). Most of the apps that did allow handwritten input were usually interactive graphing tools. For example, FluidMath can translate a user’s handwritten input into an equation, then sketch a plot of the graph of that equation, including multiple “objects” on the same plot. Figure 5 shows a sketched ellipse, line, and parabola that were translated into a single graph, with the simultaneous presentation of their algebraic expressions. When the user changes the location or shape of the objects on the coordinate plane, the algebraic expression changes accordingly and vice versa.
In sum, the richness of interactions we found in the mathematics apps is far greater than that of typical educational assessments, including the latest technology-enhanced items. Not only do tablets afford more varieties of item presentations and responses (e.g., intuitive gesture commands, accelerometer functionality, potential for handwritten responses, ability to include drawings to accompany text responses), they also create opportunities to collect data on response processes that are typically not present in traditional assessments (e.g., recording the order in which a diagram is drawn, attempts/resets on a problem before submitting the answer).

**Scoring and adaptability**

Scoring of student performance is a central component of assessment development. Scores can be used not only to evaluate student performance, but also to provide real-time feedback for students in various ways. Lower scores on a particular topic indicate more difficulty experienced and the need for additional work on that topic. Scores may also be used to determine what items may best fit a student’s current level of competence so as to keep the student engaged during assessments.
Scoring

Most apps mark a response as either right or wrong, and use percent correct as an indicator of user’s overall performance. Speed may also be considered in scoring. For example, in Math Champ Challenge, the points one may receive on an item decrease with time. Item difficulty is typically not considered in scoring, so that all items are weighted equally in calculating one’s final score. Scores in these apps are only meaningful within the apps themselves; no apps in the sample we reviewed claim any external validity of their scores, either as evidence of competency or as a basis for decisions to be made outside of the virtual world of the app. This is a critical difference from educational assessments.

Adaptability

Not all applications adapt to user responses. The ones that do typically use one of three ways to adapt the difficulty of items to user’s ability: The first approach is to let the user select a difficulty level. For example, Middle School Math HD by Interactive Elementary allows a user to pick between one of the three difficulty levels: easy, medium, and hard, before the user starts the game play. A second approach is to offer items of increasing difficulty, while successful completion of an “episode” is required to unlock the next one (e.g., Fraction Planet by Playpower labs). The last approach involves the dynamic presentation of questions. Users still need to finish the current task before they can unlock the next task, but which task will be presented next is determined by their performance on the current task (e.g., Woot Math).

In addition to the above notes, we also observed various usability issues in the mathematics apps. Instead of enumerating them here, we shall discuss ways to avoid them in the context of assessment design in the next Section.

Discussion

We conducted this survey of mathematics apps with the intent to learn what we should (and should not) do in developing innovative mathematics assessments on tablet devices. While the landscape of mathematics apps may seem vast, there are only a few categories of gameplay, interaction, and adaptability that are currently at play. These features may influence the how students engage in problem solving and learning. They also have an impact on how we can interpret students’ interactions with the interface as well as their final answers. These are critical concerns as we design the next generation tablet-based mathematics assessments. In what follows we will discuss some of the lessons learned and how they may apply to tablet-based assessment design. Many of these recommendations may apply to all Technology Based Assessment (TBA) but we leave the recommendations as specific to tablets since the interactive nature of the tablet interface is fundamentally different from a computer and students have been shown to interact with tablets and computers in different ways and with different results (see Abrams, Davoli, Du, Knapp, & Paull, 2008; Davoli, Du, Montana, Garverick, & Abrams, 2010; Davoli & Brockmole, 2012).

Mathematical content representations

By design or by accident, many mathematics apps we reviewed have inadequacies in representing the accuracy and richness of the mathematics content. While this may be forgiven in games and informal learning environments, such flaws may create systematic errors in scoring and interpretations when it comes to assessments. Fortunately, in the educational assessment field, there is a rigorous review process in place to guarantee the content accuracy, which works well for familiar item types.

One challenge we are likely to grapple with, though, is how to continue this strong tradition as assessment tasks become increasingly complex, interactive, and game-like. We observed that in some cases when a mathematics problem is operationalized in a game-like setting, where mathematical concepts map to manipulatable objects on the tablet, valid mathematical operations do not always translate to legitimate physical operations on the tablets, such as the issue described earlier in DragonBox, where the allowed actions preclude obtaining one of two valid answers. This may not be a concern for gameplay, but it could undermine the validity of an assessment. Thus one lesson learned is: In creating an interactive game-like assessment item, be very careful about how mathematical concepts...
and operations map to objects and actions in the virtual world; thoroughly review the mapping early in the task design stage.

Interactive items

One indisputable attraction of these apps is that they provide a wide range of rich and engaging interactions around mathematics content. This is what users – students, parents, and teachers – come to expect of an “iPad experience.” This, undoubtedly, is also the expectation they will bring to tablet-based assessments. The bar is set very high, and the assessment industry has much ground to cover to catch up with the apps. However, at the same time, the level of rigor needed is much higher for assessments than low/no-stakes games and certain liberties taken with content for the sake of interactivity is not acceptable for high quality assessments.

We take high quality interactive item types as a given for any serious attempts to create tablet-based mathematics assessments; they are what students have come to expect of all things tablet, and, when done right, they engage students in mathematical thinking. We nevertheless advise against creating interactive items only for the sake of engagement. As we pointed out earlier, without a clear focus and a meticulous execution, it is easy to lose sight of the construct and end up developing a fun but uninterpretable distraction.

From an assessment point of view, the biggest value of interactive tasks is that they provide data on the intermediate steps and strategies students use to solve problems. It is often hard to infer why a student answered an item incorrectly based on the final answer alone. Having a complete record of the problem-solving process helps in determining the cause of the error. Such information can help to interpret student performance in an assessment, or to determine scaffolding or instructions one might need. We provide three illustrations:

- **Final products and response processes.** Most applications use students’ answering action to calculate scores. Besides accuracy and speed, students’ change of answers can also shed light on their strategy use and knowledge. For example, higher-ability students are less likely to change their answers, although they are more likely to benefit more from their answer change (McMorris et al., 1991). Recording the sequence of answering action may also reveal the context in which students change their answers, which may help identifying the cognitive process involved in such changes. For example, changing answers on previous items after answering later questions may suggest students used information from later questions to update their thoughts. Changing answers while reviewing before submission, on the other hand, may suggest they worked the problem for a second time and identified an error.

- **Navigation through the task.** Navigation through a task can also inform us about preexisting knowledge. For example, a student may skip a question he or she is uncertain with at first, coming back to the question later. Studies on test navigation are scarce, but current evidence indicates that high-ability students are less likely to go back and forth while completing the assessment (Kim et al., 2012).

  It is worth noting that task navigation reflects not only student task processing, but also the structure of the assessment. For example, many formative assessments allow students to go back to previous items and ahead to future items via arrows from the current screen. Some apps, however, allow student to go to any previously-solved item from a review board (e.g., Woot Math, see Figure 6). Design of assessment flow will influence (a) whether students can skip a problem and come back later, (b) whether students can engage in self-monitoring actions such as marking an answer as unconfident or problematic, and (c) how students would navigate back to a previous item (see Pan, Cayton-Hodges, & Feng, 2014).

- **Problem solving and tool usage.** Problem solving processes can be revealed through students’ use of virtual tools offered in the app. Virtual tools may include calculators, hints, virtual sketch paper, and other tools. The observable problem-solving behavior may include deciding whether they want to use a tool to help solve the problem (e.g., “I want to use sketch paper” vs. “It’s easier to just do it in my head”; see, for example, Wilson, 2002, for a discussion on the off-loading of cognitive work), which tool they choose to use (e.g., “I used the calculator to do the math, but the number I got didn’t look right. I think I am going to do it by hand”), and what they do with these tools (e.g., doing long division on the virtual sketch paper vs. creating a visual model for the problem).
Interpretation of students’ interactions with tools should happen on the item level, but the aggregated measures of such interactions can also provide insights into students’ problem solving. For example, Kim et al., (2012) examined students’ writing sequence in a tablet based mathematics assessment. They reported correlation between writing pattern across items and students’ mathematics score. High-scoring students had fewer erased strokes than their peers. They also displayed more top-down writing movement in their problem solving. Researchers explained that multiple erased strokes at the same location might indicate a trial-and-error approach in students’ problem solving, while the top-down writing manner may reveal a systematic planning and problem-solving process.

At the item level, observation of students’ problem-solving behavior can provide details about students’ specific strategy use as well as their modeling with mathematics. For instance, we conducted usability studies with a prototype tablet assessment for rational number understanding. In this task, students can activate a virtual cutting board where they cut and share banana bread among plates. To solve the following question (Figure 7), two students both activated the cutting board.

Their answers to the multiple-choice question were identical. Their written explanations were also similar, as were the final screenshots when they left the cutting board. Yet they adopted different strategies. While asking for their plan about how to use the cutting board, the first student said,

“I thought about cut a loaf into 5 pieces. Then I think I might have more pieces but each piece is smaller. So I just cut each into 4 (as I did for the library) and see what happens. Each person get 3, and there’s one piece left to share between them. So each one get 3 and a tiny bit more” [sic]

The second student was also prompted about what he planned to do when he activated the cutting board:

“Four doesn’t go into five. So I want to cut each into two and try. It doesn’t work. I’ll try three. Still not working. I’ll try [to cut each into four]. Still not working... Wait now this is the same as the library! But there is a leftover. Someone can take the leftover, they will get more.” [sic]

These two strategies were representative of many of the students interviewed. In this case, we were able to map the number of exploratory actions (i.e. how many times the students reset the cutting board, recovered from interviewers’ notes and video recordings) with the students’ strategy use. In large-scale field testing, however, it is unrealistic to record all the interactions student have with the tablet. It therefore becomes crucial to pre-define events of interest among multiple possibilities in the assessment design and to iteratively adjust these captured events throughout the development and testing process to ensure that they are gathering the evidence that was intended.

To summarize the lesson on interactive items: When designing an interactive item, start with what evidence is needed to make inferences about student performance, and design the interactions to collect the necessary data.
Developing innovative item types

Our review also suggests areas of weaknesses in the design of popular mathematics apps that will require innovations in designing tablet-based assessments. We focus here on two issues, namely (a) student reflections and explanations, and (b) scaffolding.

- **Self-explanation and reflection.** Literature in cognitive psychology (Chi, De Leeuw, Chiu, & Lavancher, 1994; Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001), education, and assessment (Orsmond, Merry & Reiling, 2002; Orsmond, Merry, & Callaghan, 2004) point to the importance of self-explanation and reflection. Very few of the apps reviewed offer the opportunity for users to reflect and self-explain, however. This may be a byproduct of the game design nature of the apps (i.e., having the user’s goals be to gain points or to complete the game as quickly as possible). In education and assessment, where so much can be learned by listening to a student’s explanation (such as the example above from bread-cutting), it would be amiss if we did not provide the opportunity for reflection when feasible.

Prompts for verbal or written explanation can engage students in a process different from their initial reaction. These prompts sometimes act as a re-examination to avoid careless errors, while they sometimes engage a new strategy or modeling. In either case, explanation engages a deeper processing of the information, provides validity evidence for scoring, and can provide teachers enriched interpretation of students’ scores. Moreover, the self-explanation process is also a recommended practice for students in promoting conceptual understanding and learning transfer (Rittle-Johnson, 2006; Berthold, Eysink & Renkl, 2009).

However, prompting explanations for all items requires additional time for assessments, and providing written explanations can be especially tiring for younger students with limited keyboarding skills. These difficulties raise two questions. One is when should we prompt for explanation; the other is how should we collect these explanations. Timing of the explanation prompt can be item based or response based. Item-based prompts should be used in items where multiple steps are needed to solve the problem, as well as when multiple strategies are present in solving a problem. Response-based prompts may be used when students provide an incorrect answer, to differentiate a careless error from more serious misunderstanding. Such prompts may also be used when the student has spent too much time on a question to differentiate between a reading comprehension problem, confusion and lack of knowledge, and poor planning.

With regards to ways to collect student explanations, tablet devices offer a number of possibilities, such as using the soft keyboard or an external keyboard, handwriting, drawing, and speech input. Our pilot study (Cayton-

---

**Figure 7.** Example item (Cayton-Hodges, Feng, Pan, & Vezzu, 2013)
Hodges, Feng, Pan, & Vezzu, 2013) showed that fourth-grade students liked writing numerical responses on the tablet by hand, but preferred typing their short-response answers over writing on the tablet (although they also appreciated the ability to supplement their written explanation with drawings, which is typically difficult on a pc with a keyboard-mouse interface). Besides collecting responses through a multi-touch screen, verbal input may also be utilized. For example, Siri is available on iPad as a real-time verbal input processor. It can be called through the assessment interface and return recognized strings in real time.

Reasoning and argumentation are practices called for by the Common Core State Standards (see the Standards for Mathematical Practice, Common Core Standards Initiative, 2010), but these self-explanation skills are often downplayed in practice. There is much room for innovation in this area, and tablets provide a very versatile platform for capturing explanations and reflections, in speech, writing, or drawing. Scoring such evidence is out of the scope of this paper, though a number of natural language processing tools may be used to automatically categorize student explanations. We suggest the following: As part of the design of engaging, interactive tasks, also create opportunities for students to self-reflect or explain their problem-solving process. Consider the use of appropriate modality to capture such evidence.

- Providing aids and scaffolds. Tutorials, hints, scaffolds, and other tools are widely used in the mathematics apps we reviewed. They are done in such a way to quickly orient the user to the task and to guide the user along the path through challenges. This type of aid is particularly important for game-like apps, which risk losing users if they cannot jump into the game quickly or get stuck at a level and cannot find a way out. These tools are typically presented on an as-needed basis. Part of the reason is the limited screen real estate – apps need to maximize the area of useful interactions and avoid clustering. Hence, when and how to provide assistance to the user becomes a key consideration in app design.

In comparison, in traditional assessments the pressure is on students to follow directions and to figure things out. Copying the practice of paper-and-pencil tests, computer-based assessments today typically print instructions statically on the screen, regardless of whether the user needs them. If an assessment task is long and complicated, students often have to read one or more pages of directions. Few apps can afford to do this; they must minimize the effort and working memory load of the user, or else they lose out.

Extensive reading does take a toll on students. Through our tablet piloting, we found that students are often not very careful while reading the items, and they sometimes get confused about the demands of the item (see Cayton-Hodges, Feng, Pan, & Vezzu, 2013). Because the directions are static, students get no feedback or opportunities to confirm or challenge their understanding. In such cases, students often turn to the interviewer for help. Some students commented that they would be helpless in a summative assessment because they would not have anyone to ask.

The problem here is how we deliver “directions.” In games and apps, as well as in daily communication, helpful directions are given on an as-needed basis. Directions in testing are an anomaly, in that being standardized is a necessity. While this makes sense when item types are simple and familiar, the individualized helpfulness will become increasingly important as tasks become more complex and creative, particularly in formative assessment settings. Letting students struggle in frustration not only goes against the “iPad experience” but also threatens the validity of the assessment; a score is not a proper reflection of mathematical ability if the student is confused about task requirements. If we decide to engage students in complex assessment tasks, then it is our responsibility to ensure that all students understand and can navigate the tasks. The mathematics apps we reviewed provide many positive examples to guide students through challenges, through animated tutorials, adaptive instructions, and options to request help.

Help-seeking behaviors can also provide valuable assessment information. In a formative assessment, the teacher may want to know which students clicked on the “help” button and under what circumstances. Having context-dependent assistance in this case creates assessment opportunities. Imagine an interactive instruction (or a computer agent) that can read, rephrase, or explain parts of the directions, depending on users’ needs. This would be a helpful tool for English language learners and students with special needs, and would enhance the validity of the assessment as a result. Finally, more sophisticated aids may provide students with learning resources, from “cheat sheets” to instructional videos (see Math by YourTeacher.com). This creates opportunities to assess not only what students know (and do not know) but also how much and how quickly they can learn.
While not every method of aid is appropriate for all assessments, we recommend the following: Assessment developers need to adopt the mindset of app developers: it is the designers’ responsibility to keep the user engaged, on task, and moving forward. A test is only valid to the extent that students are “in the game” However, take caution that the construct of measure is never sacrificed in the name of engagement.

Summary

The aim of this report was to summarize a survey of mathematics education apps in the Apple App Store, conducted as part of a research project to develop a tablet-based assessment prototype for elementary mathematics. This survey was performed with the goal of understanding the design principles and techniques used in mathematics apps designed for tablets. We focused our reviews on four areas, (1) the quality of mathematical content, (2) feedback and scaffolding, (3) richness of interactions, and (4) adaptability of the applications. These four areas were cultivated from prior research on digital tools in mathematics (e.g., Digital Tools for Algebra Education criteria; Bokhove & Drijvers, 2011), designing principles of learning objects (e.g., Learning Object Evaluation Metric; Kay & Knaack, 2008), as well as quality of mathematics instruction (e.g., Hill et al., 2008). This review culminates in the formulation of four recommendations for researchers and assessment developers on designing tablet-based mathematics assessments: (1) Thoroughly review the mapping between concepts/operations and objects/actions early in the task design stage; (2) Start with what evidence is needed to make inferences about student performance, and design the interactions to collect the necessary data; (3) Create opportunities for students to self-reflect or explain their problem-solving process; and (4) Adopt the mindset of app developers to keep the user engaged, on task, and moving forward to ensure that students are “in the game” enough to accurately assess content knowledge.

References


18


Appendix

List of Apps reviewed


Using Trialogues to Measure English Language Skills

Youngsoon So1*, Diego Zapata-Rivera2, Yeonsuk Cho2, Christine Luce2 and Laura Battistini2

1Department of English Language Education, Seoul National University, Seoul, South Korea  //  2Educational Testing Service, Princeton, New Jersey, USA // yyoungsoon_so@snu.ac.kr // dzapata@ets.org // ycho@ets.org // cluce@ets.org // lbattistini@ets.org

*Corresponding author

ABSTRACT

We explored the use of technology-assisted, trialogue-based tasks to measure the English language proficiency of students learning English as a second or foreign language. A presumed benefit of the system for language assessment is its suitability for use in scenario-based tasks that integrate multiple language skills. This integration allows test developers to simulate real-life language use situations by administering more authentic test tasks to strengthen the link between test performance and its interpretation as an indicator of expected success in performing real-life language use tasks. We conducted a usability study with 20 English language learners (ELLs) in grades 3 to 5, representing ten different first languages. Results suggested that the trialogue-based tasks engaged students in an authentic experience and provided evidence of their English proficiency. Perception data indicated that students perceived the trialogue-based tasks positively and understood how to respond to the virtual characters. Implications are discussed for future research and the potential of using trialogue-based assessment tasks for measuring the proficiency of young ELLs.

Keywords

Trialogues, Computer-assisted conversations, English language learners, English as a second or foreign language, Young children

Introduction

Conversations between a student and virtual characters have been used to facilitate learning by supporting human-like interactions with computer-based systems (Chan & Baskin, 1990; Graesser, Person, Harter, & Tutoring Research Group, 2001; Johnson, Rickel, & Lester, 2000). To date, this type of system has been used for non-assessment purposes, mainly for the tutoring of cognitive skills, such as scientific reasoning (Graesser et al., 2010). Our study examined the potential of applying such a conversation-based system to assess constructs that are different from those previously implemented for non-assessment purposes. More specifically, our study explored the use of conversation-based tasks as a means of gathering evidence of the English language proficiency of students learning English as a second or foreign language (ESL/EFL).

Trialogues, interactive conversations between one student and two virtual characters, have been used to create engaging, realistic environments in which a student is positioned to possess more or less information and/or knowledge relative to each of the two virtual characters. Graesser et al. (2010) utilized trialogues between a student and two virtual characters (a peer and a teacher) to identify evidence of student inquiry skills. In this particular setting, the student is expected to help the less knowledgeable virtual peer while receiving feedback or scaffolding from the more knowledgeable virtual teacher. Students can assume different roles within trialogue-based tasks depending on their expected prior knowledge, playing the roles of either the giver or the receiver of information (Cai et al., 2009).

We focus in this study on four potential advantages of applying the trialogue system to measure the English language proficiency (ELP) of young English language learners (ELLs). First, trialogue-based interactions take place in virtual environments. Such environments are familiar to today’s young students, who are more immersed in new technologies than any previous generation (Kaiser Family Foundation, 2010). Well-designed virtual environments can involve young students in language-use situations that maximize meaningfulness, interactivity and engagement — qualities considered very important in teaching and assessing young language learners (Cameron, 2003; Hasselgren, 2000, 2005; McKay, 2005, 2006). Second, the trialogue system allows for the creation of virtual communication situations in which language learners must use language to perform the given tasks. Integration of obligatory communication situations into task design is not a new concept in second language teaching and learning. For example, information gap activities, which are commonly-used language teaching activities, deliberately
manipulate the information available to each student as a means to promote more genuine communicative interactions among language learners (Johnson, 1982; Widdowson, 1978). Third, the trialogue system can incorporate scaffolds or supports that help students demonstrate the best of their abilities within assessment tasks in a way that students perceive the scaffolds to be a natural part of their conversations with virtual characters.

Last, relative to traditional language assessment tasks, the trialogue system’s virtual environment allows for the creation of authentic scenarios in which students use language to achieve communication goals that are similar to those which they are expected to achieve in non-testing situations. The last point is particularly important when considering the purpose of language testing, i.e., to obtain evidence of competence that can be extrapolated to non-test, real-life contexts (Bachman & Palmer, 2010). Therefore, assessment tasks that are similar to the language-use tasks expected to occur in non-test situations are more likely to lead to the valid interpretation and use of test results (Kane, 2006).

In the language assessment literature, it has been argued that integrated assessment tasks — tasks that require a student to engage multiple language skills, such as listening and speaking — are a more proximal simulation of non-test language use (e.g., Lewkowicz, 1997). To successfully participate in a conversation, one must perform the roles of both listener and speaker. However, current language assessments tend to measure listening, reading, speaking and writing in isolation, and the tasks in such assessments cannot easily implement language skill integration. This limitation is partially due to the lack of practical ways to create a test in which a student’s response to a task can be used as input for the next task.

In light of the anticipated benefits of using trialogues to measure young ELLs’ English language proficiency, the purpose of our project was to develop trialogue-based tasks in which 8- to 11-year-old ELLs could achieve communication goals while engaging multiple integrated language skills. We first report on the development process and describe the resulting tasks. Next, the results of a usability study conducted in March 2013 are discussed. We conclude with implications for language assessment and suggestions for future research.

Development process

The process used in the development of the trialogue-based tasks is summarized in Figure 1. The process began with identifying the language constructs of interest, followed by developing scenarios that required students to perform communication activities that engage the identified constructs. The developed scenarios were then implemented using a dialogue authoring tool. The implemented scenarios were combined with graphical components using the Unity system (Unity Technologies, 2013) to produce the tasks evaluated in the usability study. The development generally proceeded in the order presented in Figure 1, although considerations in a later stage made us revisit earlier stages. In the sections that follow, we show how each component of the development process was operationalized.

Construct identification and scenario development

The target constructs were defined with reference to the learning objectives and communication goals commonly found in English curricula for elementary aged students in several countries where English is taught as a second or
foreign language, including Brazil, China, Korea, Japan, Mexico, and the Philippines. Table 1 summarizes the selected constructs by language skill.

<table>
<thead>
<tr>
<th>Language skill</th>
<th>Target constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening</td>
<td>• Understanding directions</td>
</tr>
<tr>
<td></td>
<td>• Understanding simple questions</td>
</tr>
<tr>
<td>Speaking</td>
<td>• Answering simple questions</td>
</tr>
<tr>
<td></td>
<td>• Identifying and summarizing key ideas</td>
</tr>
<tr>
<td></td>
<td>• Forming an appropriate request</td>
</tr>
<tr>
<td>Reading</td>
<td>• Understanding written instructions</td>
</tr>
</tbody>
</table>

After selecting the constructs to be measured, the language use contexts in which students would be expected to interact with virtual characters while engaging the targeted language constructs were created. School settings (classroom and school library) were selected for the design of the language use contexts, and five triologue-based tasks were developed, with each task designed to measure a combination of the constructs in Table 1. The tasks were designed such that different skills (e.g., understanding directions and answering simple questions) were integrated. Each task was designed to accommodate multiple possible conversation patterns between a student and the virtual characters, based on natural language processing of the student’s response. All of these conversational possibilities are visually represented as “trialogue diagrams” (Figure 2).

As illustrated in Figure 2, each task started with an opening that set the stage for interactions and ended with a closing that concluded the scenario and transitioned to the next task. The opening for each task was usually a question directed to the student. A range of possible student responses were considered and categorized as (a) correct, (b) partially correct, (c) relevant but incorrect, (d) irrelevant, or (e) no response. Each of these response categories was designed to trigger an appropriate next response from one of the virtual characters. For example, a correct response typically transitions to a closing statement, while a partially correct or incorrect response triggers a scaffold that provides the student with an opportunity to either elaborate on or correct the previous response. An irrelevant
response, or no response, triggers the repetition of the opening question. Each anticipated conversational pattern diagramed in Figure 2 was designed to provide evidence of the student’s level of ability related to the target constructs.

In the specific task illustrated in Figure 2, three constructs were targeted: understanding directions, understanding simple questions, and answering simple questions. The first target construct was measured by examining the presence and amount of important information from the directions that were included in a student’s response(s). For example, if a student provided a partially correct response in the first attempt but was unable to elaborate on the original response after a follow-up scaffold question was asked, that would be taken as evidence that the student did not fully understand the spoken directions provided and, therefore, possessed lower ability on the construct of understanding directions. The latter two constructs, understanding simple questions and answering simple questions, were measured by observing how well a student sustained the conversation with the virtual characters, as evidence of the student’s understanding and appropriately responding to their questions. Thus, the fact that the student in the example above did not respond appropriately to the scaffold question would be taken as evidence that he or she did not understand the intention of the second question and/or did not know how to respond. This type of response pattern provided evidence that the student did not demonstrate high ability on the constructs of understanding simple questions and answering simple questions.

Scenario implementation and graphics integration

The scenarios developed in the earlier stage were implemented into computer-based conversations using a dialogue authoring tool based on AutoTutor Script Authoring Tool (Susarla, Adcock, van Eck, Moreno, & Graesser, 2003). Anticipated student responses were tested, and necessary revisions were made to ensure that the conversations between the student and the virtual characters proceeded as anticipated. The example in Table 2 illustrates one of the many conversation flows shown in Figure 2. The conversation starts with the virtual character Ron asking the student What are we learning about today?, to which the student provides the partially correct response Weather, which is then followed by scaffolds from the virtual characters. In this instance, the virtual character Lisa delivers the scaffold: Yes, but it’s not about any weather. You need to tell Ron more. Lisa’s instruction then triggers Ron to rephrase his original question: What are we learning about the weather? The student is then given the opportunity to elaborate upon the previous answer, saying Weather around the world. Finally, Lisa closes the conversation with an acknowledgement of the correctness of the response.

<table>
<thead>
<tr>
<th>Character</th>
<th>Utterance</th>
<th>Character</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ron:</td>
<td>What are we learning about today?</td>
<td>Student:</td>
<td>Weather.</td>
</tr>
<tr>
<td>Lisa:</td>
<td>Yes, but it’s not about any weather. You need to tell Ron more.</td>
<td>Ron:</td>
<td>What are we learning about the weather?</td>
</tr>
<tr>
<td>Student:</td>
<td>Weather around the world.</td>
<td>Lisa:</td>
<td>That’s right. We’re learning about the weather around the world.</td>
</tr>
</tbody>
</table>

The implemented conversations were integrated with graphics and sound files to produce the final tasks. The audio statements and questions produced by each virtual character were recorded using text-to-speech software and were incorporated into the tasks in audio format only; the written text was not displayed for the students. This decision was based on two related considerations. First, the real-life language use situations that the trialogue-based tasks are designed to approximate take place as oral communication. Second, measuring listening proficiency, in isolation from reading ability, is necessary to estimate a student’s ability to participate in such conversations. By providing students with only the virtual characters’ oral speech and excluding written text, a representation of the students’ ability to perform non-test tasks can be inferred more accurately from their test task performance. Thus, written language was included in the tasks only when reading comprehension was part of the construct to be measured, as shown in Figure 3.

In the scenario shown in Figure 3, the student is expected to discuss one of the library rules with a virtual character (Ron) on the basis of understanding what is written on the board. A blinking arrow was added to the left side of the screen to draw the student’s attention to the board. In this task, understanding written instructions (reading) is the
main construct measured, and two additional constructs, understanding simple questions (listening) and answering simple questions (speaking), are also measured through the conversation between a student and the virtual characters. It should be noted, however, that an automated speech recognition (ASR) engine had not yet been incorporated into the tasks at the time of the usability study reported in this paper. Because the technological capability to automatically recognize and accurately respond to students’ spoken responses was not yet available, students were asked to provide two responses to each task: one spoken and one written. When students were directed by a virtual character to respond, they were first asked to record their spoken responses, and then to type what they said in a space that was provided. The spoken responses served two purposes: (a) to provide students with a more natural feeling of participating in an oral conversation and (b) to collect speech samples to train an ASR engine that would be used at a later stage of development in the project, discussed later in this paper. Thus, in the tasks used for data collection in 2013, the trialogue system used the students’ written responses to determine the virtual character’s next responses.

Evaluation: Usability study

The efforts to develop trialogue-based tasks were planned as a multiyear project with both long-term goals and short-term milestones. In the usability study, conducted in March 2013, we collected preliminary evidence on the usefulness of the trialogue-based tasks as a means to measure target constructs, with the following three objectives:

- to examine whether students tend to follow the “conversational flow” (see Figure 2) as expected, and if not, determine what modifications should be made to the current trialogue diagrams to better simulate how people actually speak in real life
- to gather lessons about the factors to be considered in developing additional trialogue-based tasks for the future
- to collect speech samples to be used to train an ASR engine for a future interactive speaking test

The first two objectives reflected short-term milestones, and the last objective was related to a more far-reaching, long-term goal to make the existing trialogue-based tasks into actual speaking tasks capable of supporting a spoken dialogue between a student and virtual characters. Therefore, the research questions of the study focused on collecting evidence related to the first two objectives:
RQ1: Are the triologue-based tasks usable for ELLs in grades 3 to 5? Can students in the target population interact with the triologue-based tasks?

RQ2: Can we gather useful evidence of particular English language constructs using triologue-based tasks?

RQ3: How engaging for students are the triologue-based tasks compared to multiple-choice questions that target the same English language constructs?

Participants

Twenty students in grades 3 to 5, classified as ELLs by state criteria, were recruited from four public schools in a northeastern state in the United States. Student background information on gender, grade, first language, and English proficiency is summarized in Table 3. The proficiency levels were reported by an ESL teacher in each school; therefore, these teacher-reported proficiency levels may not be directly comparable across schools. However, they were useful as indicators of students’ English language proficiency. The students represented a diverse range of first language backgrounds, with 10 first languages represented among the 20 participants. This criterion was considered critical in recruitment to ensure that the speech samples to be collected would have sufficient diversity to train an ASR engine at a later stage of the project.

<table>
<thead>
<tr>
<th>Gender</th>
<th>First Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Arabic 1</td>
</tr>
<tr>
<td>Male</td>
<td>Chinese 3</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>French 1</td>
</tr>
<tr>
<td>4th</td>
<td>Hindi 3</td>
</tr>
<tr>
<td>5th</td>
<td>Korean 2</td>
</tr>
<tr>
<td>English proficiency</td>
<td>Japanese 4</td>
</tr>
<tr>
<td>Beginning</td>
<td>Norwegian 1</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Spanish 2</td>
</tr>
<tr>
<td>Advanced</td>
<td>Telugu 2</td>
</tr>
</tbody>
</table>

Materials

The following materials were used in the usability study.

Trialogue-based tasks: The tasks described in the Development process section of this paper were administered to each individual student.

Multiple-choice questions: Multiple-choice (MC) questions were chosen from an international standardized test that targets similar constructs to those targeted by the triologue-based tasks: understanding short spoken discourse, understanding short written text, and understanding short questions. Ten listening comprehension questions and eight reading comprehension questions were selected. Although these questions are normally delivered in paper-and-pencil format for operational administration, the format was modified to be delivered via computer for the purposes of this study. The decision to perform this modification was made to control the variable of delivery medium between the triologue-based tasks and the MC questions.

Background questionnaire: A background questionnaire was designed to gather participants’ personal information (gender, age, and first language), English language learning experience, and computer use.

Usability questionnaire: Usability questions were used to investigate the clarity of the directions, the appropriateness of the time given to complete each task, the perceived difficulty of the tasks, the appropriateness of the speed of the listening input, and the overall experience of participants while they completed each task.

Table 3. Background information of the participating students
Engagement questions: Engagement questions investigated the participants’ overall experiences with both types of tasks (i.e., trialogue-based and MC questions). The main focus was to examine whether participants perceived any differences between taking the trialogue-based tasks and the MC questions.

Procedure

Participating students were recruited with parental consent, and parents were asked to fill out the student background questionnaire on behalf of their children.

Over a two-week period, students participated in individual study sessions with two researchers per student, with one researcher acting as facilitator and administering the assessment tasks and the other acting as observer and taking notes on the students’ behaviors and reactions. A typical study session lasted about 60 minutes and included the trialogue-based tasks, MC questions, and follow-up cognitive interviews using the usability and engagement questionnaires (see the Materials section). The order of the trialogue-based tasks and the MC questions was counterbalanced so that 10 participants were administered the trialogue-based tasks first and the other 10 students received the MC questions first. All study sessions took place at the students’ schools and were recorded. Usability and engagement questions were administered immediately following each type of task, and engagement questions requiring comparison of the two types of tasks were administered at the end of the study session.

Results

Student responses and researcher observation notes were gathered during the 20 individual cognitive interview sessions. Data were analyzed to evaluate the evidence collected on the students’ English language skills and task engagement. In this section, the results of the collected data are discussed in relation to each of the three research questions addressed in the study.

RQ1: Are the trialogue-based tasks usable for ELLs in grades 3 to 5? Can students in the target population interact with the trialogue-based tasks?

This research question addressed the utility of the trialogue-based tasks for the intended student population by examining to what extent the participating students’ familiarity with technology was related to their reactions to the trialogue-based tasks. In order to first investigate students’ familiarity with technology, students were asked questions about their use of computers, either at home or at school, for four different purposes listed in Table 4. Results from these questions were later compared with students’ reactions to the trialogue-based tasks to investigate any relationship. Table 4 summarizes the student responses.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>More than once a week</th>
<th>Once a week</th>
<th>Once a month</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do homework</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Play games</td>
<td>5</td>
<td>12</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Use the Internet</td>
<td>11</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Watch a movie, video, or DVD</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

With the exception of one student, an eight-year-old third grader who answered that he did not use a computer for any purpose (ID#19), the students all reported that they used a computer for at least one of the four purposes. Because the participants were familiar with computers to a certain extent, all of them were able to determine what they were expected to do to complete the trialogue-based tasks. Even when students did not immediately understand what to do, they eventually understood the task without any explicit assistance from the facilitator. All 20 participants completed the trialogue-based tasks. This fact provides evidence that the target student population is likely to be able to respond to the tasks used in the study without great difficulty in understanding how to complete the tasks.

RQ2: Can we gather useful evidence of particular English language constructs using trialogue-based tasks?
This research question addressed whether the trialogue-based tasks gathered useful evidence of ability related to particular English language constructs. The question was examined from two perspectives. First, the students’ performances on the trialogue-based tasks were compared with their levels of English language proficiency as identified either by their teachers or by their performances on the MC questions. Second, the same research question was indirectly addressed by investigating the extent to which the students engaged with the communication tasks as intended, with minimal influence by construct-irrelevant factors.

The collected evidence indicated that a positive association existed between students’ English language proficiency and their performance on the trialogue-based tasks. First, three students who were identified as advanced ELLs by their teachers (ID#4, #10, and #14) performed very well on the trialogue-based tasks. They either answered all the questions correctly on their first attempt or were able to elaborate on their partially correct responses when the virtual characters provided scaffolds. On the other hand, six students who scored below average on the MC questions (ID#2, #3, #7, #8, #19, and #20) did not perform as highly on the trialogue-based tasks. For example, the student who scored the lowest on the MC questions (ID#2) could neither provide correct responses on the first attempt to any of the opening questions in the trialogue-based tasks nor correct his previous responses when given scaffolds. This tendency to not attempt to modify one’s previous responses, even in the presence of scaffolded feedback, was found across the six students with lower scores on the MC questions. These two observations constitute preliminary evidence that trialogue-based tasks can provide evidence of ELLs’ language proficiency.

In addition, responses to usability questions were analyzed to investigate whether students perceived trialogue-based tasks positively, which was assumed to be a necessary condition for the use of trialogue-based tasks in measuring English language constructs. Nine usability questions were asked at the end of the trialogue-based task session. Figure 4 summarizes the nine questions and the response frequencies.

The results generally suggested that students perceived the trialogue-based tasks positively. Participants reported that they understood how they were supposed to respond to the virtual characters (Q#1), and they had enough time to answer the questions (Q#2). Furthermore, students generally did not perceive the pace as too fast (Q#3) or think that the test was difficult (Q#5). Some particularly encouraging findings included many students reporting that they liked speaking with the virtual characters (Q#7) and that they felt that the virtual characters understood what they said (Q#8). Similarly, almost all students (19 out of 20) reported that what the characters said made sense (Q#9).
Despite these positive findings, the responses of the one student who reported not using a computer for any purpose (ID#19) were mostly negative. This student reported that he did not understand what to do in the trialogue-based tasks and that he did not feel that the virtual characters understood what he had said. This suggests that there may be a relationship between students’ familiarity with computers and their experience taking the trialogue-based tasks.

During the usability study, students were given an opportunity to openly express their opinions about the trialogue-based tasks to capture any perceptions that were not elicited by the usability questions. Some students reported that they liked the trialogue-based tasks because they wanted to practice speaking and thought they could learn English better by doing this type of task. Two students (ID#10 and #13) reported that they liked the trialogue-based tasks because they simulated real life. However, other students shared negative feedback about some aspects of the trialogue-based tasks. For example, eight students reported that in some scenarios it was difficult to remember important information needed to answer the questions, and three students reported that they found the virtual characters’ rate of speech to be too fast. In addition, three students did not like the fact that they had to speak (ID#1, #6, and #12). One student (ID#6) chose not to speak and instead only typed his responses. When asked, he reported that he did not want to speak. This finding may be attributable to the fact that some second language learners may not feel comfortable speaking in English.

Some areas for improvement emerging from the usability study will be considered in the later stages of this project in order to enhance the user experience. First, several students reported that they did not like the appearance of the virtual characters because the characters did not look realistic. Second, one student (ID#5) commented that he did not like the trialogue-based tasks because the characters kept asking him to repeat what he had already said. This comment suggests a possible limitation of the current system and/or of the conversation flows administered in the trialogue-based tasks. Despite such negative comments by a few students, the participating students were positive overall about the trialogue-based tasks. This result indicated that the trialogue-based tasks have the potential to be used as an innovative assessment tool to measure the English language proficiency of young ELLs.

**RQ1:** How engaging for students are the trialogue-based tasks compared to multiple-choice questions that target the same English language constructs?

To address the final research question, a comparison was made of students’ levels of engagement when using the trialogue-based tasks and the MC questions. Table 5 summarizes the students’ responses to the three comparison questions.

<table>
<thead>
<tr>
<th>Comparison question</th>
<th>Trialogue</th>
<th>MC questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which test did you find more difficult?</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>On which test do you think you did better?</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Which test would you choose to take?</td>
<td>7</td>
<td>12</td>
</tr>
</tbody>
</table>

The results in Table 5 indicate that students reported thinking that they had performed better on the MC questions, and, if given the choice, they would choose to take the MC test over the trialogue test. This was a somewhat surprising finding given that student responses to the trialogue-based tasks were generally positive. Responses to follow-up probes revealed that the main reason students preferred the MC question format was due to their greater familiarity with that format and their consequent anticipation of better performance on the MC items.

Nevertheless, it should be noted that speaking was not required in response to the MC questions, and therefore, it may not be fair to compare student engagement between the two formats. This interpretation is strengthened by the fact that some students reported liking the MC questions because no speaking was required (ID#1, #6 and #12). These students’ responses imply that second language learners may feel less confident using productive language skills (e.g., speaking) than receptive skills (listening and reading). However, there were students who preferred the trialogue-based tasks (ID#15, #16, and #18) despite the fact that their perceived level of performance was higher for the MC questions. These students reported preferring the trialogue-based tasks mainly because they felt that they could “practice more talking” and “learn English better,” or they “wanted to be good at speaking.” Another student who reported preferring the MC test for assessment (ID#13) explicitly stated that she would choose the trialogue-based tasks “if the purpose is learning, not testing.” This student also commented that the trialogue-based tasks seemed to “simulate the real life [sic].” The students’ perception that the trialogue-based tasks would help improve
skills that they personally wanted to build is discussed in the next section with regard to the implications of using such tasks in language assessment.

Discussion

The primary goal of this study was to investigate the usefulness of computer-mediated, conversation-based tasks, more specifically trialogue-based tasks, as a means of language assessment. The reported study provided initial evidence that students engaged with the interactions that were built into the trialogue-based tasks as anticipated, and that the students’ experiences with the trialogue-based tasks were generally positive. In this section, implications of the results of the usability study are discussed from the following perspectives: (a) the representation of language use tasks in non-assessment situations, (b) potential construct-irrelevant factors, (c) the collection of formative information, and (d) the provision of meaningful communication opportunities for students with limited exposure to English communication.

Overall, the results of the usability study constitute evidence supporting the use of trialogue-based tasks with young ELLs as a means of providing a more engaging test-taking experience. The students’ positive responses to the usability questions (see Figure 4) are important, because we assume that if students did not perceive the virtual characters as understanding and responding appropriately to their utterances, then the language samples elicited by the trialogue-based tasks would not bear much resemblance to non-test communicative language use. In addition, the finding that some students liked the trialogue-based tasks because the tasks “simulated real life” and, compared to MC questions, were “more realistic because the virtual characters talk” was very encouraging. When students perceive that assessment tasks are similar to the language use tasks that they perform in non-assessment contexts, it is more likely that they will engage in the abilities required for real-life situations. This finding is encouraging, as it hints at the potential of trialogue-based tasks to enhance the validity of interpretations made on the basis of student test performances. If, while completing trialogue-based language tasks, students engage in language and communication processes that are similar to those required in non-assessment situations, their assessment performances may be better indicators of their expected success in non-test, real-life communication tasks, thus providing a basis for more accurate and adequate inferences on the basis of test scores. Furthermore, the use of such tasks, being more proximal to real-life language use, may bring about a more positive washback effect in language learning and teaching by motivating language teachers to make greater use of tasks that help their students engage important language skills in more meaningful ways.

Despite these encouraging findings, several factors reported may have introduced construct-irrelevant variance in student performance on the trialogue-based tasks. For example, some students (ID#3, #6, #11, and #16) commented on the unrealistic appearance of the virtual characters. Although it is unknown to what extent the students’ attention to the appearance of the virtual characters actually influenced their performance on the tasks, these comments suggest that the appearance of the virtual characters may have influenced student engagement during task interactions. Another potentially construct-irrelevant factor was related to memory load. Some students (ID#13 and #18) reported that they had difficulty retaining the information needed to successfully perform some of the tasks. It is important to note, however, that this situation mimics real-life experience and may actually contribute to the trialogue-based tasks’ ability to capture the students’ deployment of language skills in an authentic communicative interaction. It is not unexpected that such a task may place a somewhat greater burden on memory than the MC questions. Nonetheless, these findings provided useful insights into factors that should be taken into consideration when designing trialogue-based tasks to minimize the influence of potential construct-irrelevant factors and thereby maximize the validity of the interpretation of performances on the tasks as indicators of the target constructs.

With respect to the collection of formative information via the trialogue-based tasks, we would argue that the fact that scaffolds can be easily embedded within trialogue-based tasks makes the tasks potentially very useful for formative assessment. Depending on the specific conversational route that students take, they are given opportunities to correct or elaborate upon their previous responses by making use of differing amounts of the scaffolding provided. One student (ID#4) reported that he liked the trialogue-based tasks because “the characters respond [to] a wrong response.” This example clearly illustrates that the student was able to notice, on the basis of the characters’ reactions to his response, that something was not quite right. Furthermore, it was observed that not every participating student responded to the same scaffolds in the same way, and not all students were able to utilize the additional scaffolds provided. This variability implies that student ability levels on the target constructs can be
Differentiated depending on the specific ways students interact with trialogue-based tasks (i.e., whether they provide a correct response in the first attempt and whether they benefit from the provision of additional scaffolds). This design feature is important because it allows us to gather more information on where a student’s strengths and weaknesses lie, indicating areas on which students should focus for improvement.

Finally, we are confident that the communicative tasks investigated in this study hold potential to provide meaningful English learning opportunities to students learning English in countries where exposure to English language is generally restricted to classes at school. These English classes are often taught by non-native English speaking teachers who share a first language with their students. Students in such contexts may encounter relatively limited opportunities in which they are required to use English. This point is well illustrated by comments from students (ID#15, #16, and #18), who said that they found the trialogue-based tasks useful to practice speaking in English. This finding, that even students who are learning English in an English-speaking country seek more opportunities to practice speaking, provides a strong case for the need for such learning opportunities for students learning English in countries where English is not the primary or official language. Related to this point, the data collected from the usability study showed that some students were reluctant to speak and preferred the MC questions because they did not require any spoken responses (ID#1, #6, and #12). This finding indicates that, through trialogue-based tasks, students can gain opportunities to practice the skills in which they have less confidence and need more practice. We expect such opportunities to be helpful in students’ goal of developing well-balanced language proficiency across all language skills. We also note that, in some instances, the trialogue-based tasks may hold an advantage over face-to-face communication for students who feel less intimidated when interacting with virtual characters than with real humans. This last point is particularly relevant for young students, who are the target population of the present project, as achieving positive affective reactions with this age group is thought to be essential for gleaning meaningful performances from assessment tasks (McKay, 2006).

This study had several limitations. No inferential statistical analysis was conducted to investigate the relationship of student performances on the trialogue-based tasks with other criteria of their English proficiency. The decision not to perform the statistical analysis was made on the basis of the following three reasons. First, the sample size \( (N = 20) \) was small. Second, the participating students performed very well on the MC questions (\( \text{Mean} = 17.0 \) out of 18 possible points, \( \text{SD} = 1.39 \)). Given this narrow distribution of scores on the MC questions, meaningful results from inferential statistics were not expected. The third reason for not conducting a statistical analysis was that the methods used to score responses to the trialogue-based tasks need further validation. This is due to the fact that the trialogue-based tasks are very different from traditionally used language assessment tasks, and more research is needed regarding appropriate methods to score their responses. We discuss this point further in the next section, which addresses the future direction of the research and development project.

**Conclusions and future work**

Trialogue-based tasks have the potential to be used as tools to measure the interactive nature of language use, as discussed in the present paper using the data collected from a usability study. The types of tasks designed for the trialogue system require not only that students contribute to the tasks, but also that they receive immediate feedback on their performances. The performance feedback provided by the virtual characters can also serve as scaffolding, allowing the students to demonstrate their abilities both with and without this additional support.

 Planned future work includes incorporating an ASR system to accept spoken input directly, freeing students from the need to type their responses. If an ASR system can be implemented, the trialogue-based tasks are expected to function in a more engaging and authentic manner, enhancing their usefulness as a measure of speaking ability. Subsequent to the data collection reported in this paper, an ASR system has been developed for the tasks, and a usability study was conducted with the ASR-enabled version. The results from the ASR-enabled version are being analyzed at the time of the writing of this paper, and they are expected to be disseminated in a future publication. In addition, scoring models are currently being validated to score responses collected through trialogue-based tasks. Given the more integrated nature of such tasks, scoring methods that have been used for more traditional discrete tasks may not be readily applicable to responses collected from the trialogue-based tasks. Next, additional trialogue-based tasks are also under development to measure language constructs that were not targeted in the tasks discussed in this paper. Finally, a larger-scale data collection is scheduled with an expanded version of the assessment that is ASR-enabled and includes more tasks targeting the measurement of more constructs.
Acknowledgements

The research reported in this article was conducted when the first and corresponding author was employed by Educational Testing Service. The authors would like to thank the editor and the two anonymous reviewers of Educational Technology & Society for their feedback. The authors would also like to acknowledge their external collaborators Art Graesser, Carol Forsyth, and Zhiqiang Cai from The University of Memphis, as well as Kyle Staves and his colleagues from SortaSoft, LLC, for their assistance in the instrument development. Finally, the authors acknowledge Don Powers, Irvin Katz, Alexis Lopez, Tanner Jackson, Kristin Williamson Worden, and Ian Blood for their useful comments on an earlier version of this paper.

References


M-AssIST: Interaction and Scaffolding Matters in Authentic Assessment

Patricia Santos\textsuperscript{1*}, John Cook\textsuperscript{1} and Davinia Hernández-Leo\textsuperscript{2}

\textsuperscript{1}Department of Arts and Cultural Industries, University of the West of England, Bristol, UK // \textsuperscript{2}Department of Technologies and Information Technologies, University Pompeu Fabra, Barcelona, Spain // Patricia.Santosrodriguez@uwe.ac.uk // John2.Cook@uwe.ac.uk // davinia.hernandez@upf.edu

*Corresponding author

ABSTRACT

Authentic assessment is important in formal and informal learning. Technology has the potential to be used to support the assessment of higher order skills particularly with respect to real life tasks. In particular, the use of mobile devices allows the learner to increase her interactions with physical objects, various environments (indoors and outdoors spaces), augmented digital information and with peers. Those interactions can be monitored and automatically assessed in a way that is similar to traditional objective tests. However, in order to facilitate a meaningful interaction with formative purposes, we propose that the assessment process can be assisted through scaffolding mechanisms that transform the mobile system into a ‘more capable peer’. In this context, this paper presents the m-AssIST model which captures the necessary emergent properties to design and analyse m-assessment activities. The model is used to analyse the benefits and limitations of existing m-test based systems. This paper discusses the importance of meaningful interactions, and the provision of scaffolding mechanisms to support formative and authentic assessment.

Keywords

Mobile-assessment, Authentic assessment, Meaningful interactions, Scaffolding

Introduction

Authenticity in assessment has been widely discussed by several authors in this research field. As Brown, et al., (1989) assert, authentic or situated learning can be seen from the perspective of social constructivism, it occurs when the learner takes part in relevant activities to her real life and which take place within a culture similar to an applied context. Gulikers et al., (2004) state that current educational and professional curriculums are more focused on the development of competences than on a simple knowledge acquisition. In this sense, ILA-practices (Instruction, Learning and Assessment) are characterized as following an instructional-approach focused on learning and competence development, supporting reflective-active knowledge construction, contextualized, interpretative and performance assessment (Birenbaum, 2003). The goal of ILA practices is the acquisition of higher-order thinking skills instead of factual knowledge and basic skills. As a consequence the assessment has a formative goal. However, assessment in formal learning is commonly held to contribute to feedback to students on their learning and the certification of their achievement (Boud & Falchikov, 2006).

The term “higher-order thinking skills” (HOTS) is used to delineate cognitive activities that are beyond the stage of understanding and lower levels of application (Bloom & Krathwohl, 1956). In this paper, we use the definition provided by Kaminsky et al., (1997) who define these skills as the ones where the level of thinking depends upon the context, with a real-world situation, where the individual has to apply, reorganize and embellish knowledge in the context of the thinking situation. Assessing learners’ HOTS is important in formal learning but is particularly important in informal learning contexts such as the workplace. One key proponent of an empirical tradition of work-based learning research is Eraut (2007) the author proposes a typology of early career learning, where he emphasizes the importance of consolidating, extending and refining skills in workplace environments. In this context, self-assessment has an important role in order to support staff and management to develop their skills (Boud, 2013). Considering the future of younger learners, some authors have stated that new ways of thinking about assessment in formal education have to be proposed (for example Boud, 2006; Davies, 2010). Brown et al. (1989) identifies: “many of the activities students undertake (in formal education) are simply not the activities of practitioners and would not make sense or be endorsed by the cultures to which they are attributed... When authentic activities are transferred to the classroom, their context is inevitably transmuted; they become classroom tasks and part of the school culture... Classroom tasks, therefore, can completely fail to provide the contextual features that allow authentic activity.”

In this context authentic means that learners should be able to demonstrate and practice their skills like in real life tasks. One of the main problems in formal learning occurs when some of the considered authentic skills are included
in educational curriculums or training programmes, but the methods to assess them are not appropriate (Benett, 1993; Boyle & Hutchison, 2009). The main aim of this approach is to provide meaningful learning/assessment and promoting the transfer of knowledge to real-life situations (Choi & Hannafin, 1995). Choi and Hannafin state that knowledge is situated, when it is product of an activity, context and culture in which it is developed and used. We agree with the vision proposed by these authors and for this reason this paper claims that mobile technology is the most suitable instrument for supporting the assessment of authentic/situated tasks due to its “situational” facilities.

The European Network of Excellence in Technology Enhanced Learning (TEL), STELLAR published a report with the title “Education in the wild: contextual and location-based mobile learning in action” (Brown, 2010) claiming that a re-conceptualization of educational theories have to be undertaken in order to develop an approach for the present and future of learning designs. The use of mobile devices enables technicians and practitioners to rethink learning, teaching and assessment strategies. As Engeström (1999) states: “human activity is endlessly multifaceted, mobile, and rich in variations of content and form...the theory of activity should reflect that richness and mobility.”

In related research, Cook (2010) proposes the term “Augmented Contexts for Development” (ACD) where: learners can use mobile devices (e.g., smartphones) in combination with augmented digital information to interact with each other and interact with the physical and a virtual environment with the goal of creating their own ACD. The smartphone and the location-triggered learning activity it holds are considered here as the “more capable peer” (Vygotsky, 1978, p. 86) because the device provides guidance to the learner in order to solve a problem and as a consequence the learner augments her context of development. The main elements to develop the ACD are: (a) the physical environment, (b) a pedagogical plan (e.g., an assessment activity), (c) tools/devices for an augmented oriented approach, (d) learner co-constructed “temporal context for development” and (e) collaborative learners’ interpersonal interactions using tools. We observe here the importance of interacting with the environment and with other peers, for this reason we use the definition provided by Woo and Reeves (2007) where they present the term “meaningful interaction” as those “interactions that have direct influence on learners’ intellectual growth.” The ACD approach is also supported by the definition of cognitive systems proposed by Hollan et al., (2000) which claim that the organization of cognitive systems is extended to cover the interactions between people and with the resources and materials of the environment distributed in time.

Taking into account the above approaches we propose that a pedagogical plan can consist in an authentic assessment activity performed in a real-world environment. But we claim that this plan has to be supported by scaffolding mechanisms in order to help the learner to solve the problem by doing meaningful interactions through her mobile device. We use the description of scaffolding provided by (Wood et al., 1976): “process that enables a child or a novice to solve a problem, carry out a task, or achieve a goal which would be beyond his unassisted efforts.” As Lieberman & Linn (1991) state, scaffolding is one of several ways to encourage self-directed learning in students, especially in terms of TEL scaffolding it is useful to support self-assessment and formative assessment methods and provide automatic assistance. This paper claims that the monitoring of the interactions done by the learner during the assessment activity through her mobile device can be used to assist the learner in real time and provide helpful assistance in order to achieve the adequate meaningful interactions and corresponding knowledge. Our view is that mobile devices have the potential to support more authentic types of assessment integrated within real locations/objects in ways that stimulate and motivate learners and support their meaning making.

From these approaches, we assert that: the Environment, Time, Interactions and Scaffolding mechanisms matter when designing an authentic assessment activity for a corresponding context. We use the term “emergent properties” proposed by Dourish (2004) to refer to the elements that allow us to define the context of an authentic assessment activity. In particular we show how the meditation of those interactions through mobile devices is a key issue to ensure the authenticity of the activity in a specific yet emergent context (i.e., time + location). We capture these emergent properties in a model called m-AssIST. The m-AssIST model can be used to analyse existing m-assessment scenarios in order to identify limitations related to the assessment of HOTS. The model can also be used during the design process of an assessment activity in order to reflect about the importance of these properties for the success of the authenticity of the activity.

From the wider number of different assessment instruments, this paper focuses its analysis on the use of automatic methods of assessment (i.e., objective tests commonly used to support self-assessment). We agree with the description provided by Boud (2013) that self-assessment has to “stress the importance of learners constructing rather than receiving knowledge, of promoting the taking of responsibility for learning.” So, this previous sentence
stresses the formative strength of self-assessment methods, and the importance of changing the role of the learner from a passive one to an active one through the use of mobile devices.

In the following section, we discuss concepts for assessment in TEL and scaffolding mechanisms (we define these below) to support the assessment and development of authentic tasks and HOTS. After the discussion of the literature, the emergent properties of the m-AssiSt model are described. In order to show the application of the m-AssiSt model, this paper also contains a section reviewing three existing m-assessment scenarios where the goal was to assess situated HOTS, this review will be followed by an analysis using the m-AssiSt model in order to identify those scaffolding mechanisms and strategies to support meaningful interactions which could be applied in the selected scenarios. The scenarios are significant because they have taken into consideration scaffolding strategies to support an authentic m-assessment activity. The three scenarios are sufficiently different to illustrate diverse ways of interacting with authentic environments (i.e., a natural park, an old city district, and a botanical garden) in order to assess different HOTS assigned to different levels of learning and subjects. Finally, we finish with a final reflection of the benefits of the emergent properties captured in the m-AssiSt model.

**Relevant literature**

In this section we elaborate on our claim that mobile devices have the potential to augment and support authentic assessment. In order to do that, we first need to understand how technology has been employed to support traditional assessment, and how the analysis of the limitations of traditional assessment has allowed the identification of new ways of supporting more authentic assessment in TEL.

**Computer Assisted Assessment**

Most of researchers in the Computer Assisted Assessment (CAA) field have concentrated their efforts in creating innovative question-items or using complex computational solutions to represent complex questions and tests. They claim that the creation of new types of questions can improve the measuring of learners’ HOTS. Creating advanced assessment scenarios should involve the use of technology to perform test-activities which would be impossible or very expensive to reproduce using traditional methods of assessment (e.g., paper & pencil). Computers offer the possibility of using simulations, managing a big quantity of updated and enriched information, increasing the interaction with the information and making the learner more participative in the assessment process (Conole & Warburton, 2005).

Bull & McKenna (2003) show in their blueprint for CAA that the use of objective tests (or automatic tests) provides advantages such as: improving the interactivity with the learning content, presenting a question, obtaining a response, evaluating a response, providing a mark and answering with feedback automatically. Despite the benefits of using automatic tests, the traditional transcription from paper to computer is not enough to assess authentic tasks and HOTS such as: problem solving, problem exploration, applying, collaboration, analysing, discovering rules, spatial or time perception, among others (Mayotte, 2010).

Previous research projects have studied how to use objective tests in lifelong learning contexts (considered to mean more “authentic” contexts than the classroom). On the one hand, the EU FP7 research project ‘TEN-Competence’ developed an Open Source environment to support self-directed lifelong learning based on IMS standards (Louys et al., 2009). The system included web-based tools for self-assessment and embedded-assessment in competence oriented scenarios. The self-assessment outcomes (obtained after answering an automatic test) were used to help the learner to identify her competence gaps and generate her learning path for competence development. Furthermore, the ICOPER project (Crespo et al., 2010) was focused on finding ways that can effectively bring together educational achievements that should result in outcome-based assessment and units of learning. One of the contributions of this project was a conceptual map for learning assessment. This model comprises the whole process, including the activity of the learners taking a computational exam, as well as the phase where instructors include the corresponding feedback/grades. The project showed how the adoption of IMS standards (IMS, 2006) for learning outcomes definitions facilitates to some extent the alignment between assessment outcomes and the educational curriculums. Both projects contributed significantly to the research field of CAA, however with the fast evolution of the technology in the last 5 years we find that some of the contributions are outdated for the following reason: IMS does
not consider complex representations of questions and tests, instead it is typically used to represent traditional types of questions such as Multiple Choice or Multiple Response in order to assess knowledge. As a consequence, the tests developed by these computational-based systems do not support the assessment of situated knowledge, where the assessment activity is developed at least in a similar context (i.e., location) than the real associated one.

By drawing on our analysis of the previous approaches, we can now identify the main limitation of using computers to assess authentic tasks. First of all, its staticity: although complex simulations of the real-world can be pre-defined, this kind of approach is limited in terms of being used in any situational (uncontrolled) activity where the location and specific time where the action is performed can have an impact over the solution that has to be applied by the learner. Second, it is not possible to use this approach to interact with physical objects or people in an outdoors environment like the ACD example given above.

Boyle & Hutchison (2009) stated that the more technological resources you use, the higher skills and more sophisticated tasks can be evaluated. The JISC report “Effective assessment in digital age” (Davies, 2010) claimed that technology has to be used to create authentic assessment. For this reason some researchers state that the use of handheld devices will transform the learning and assessment practices (Thompson, 2005), opening up the possibilities of using other technologies to support assessment based on tests. In particular, the wider use of mobile devices has come up with a positive transformation in the assessment of authentic practices.

Computing – Based testing: Adding authenticity to Computer Assisted Assessment

As proposed in a previous publication (Santos et al., 2012) we consider the term “Computing” instead of “Computer” as more appropriate when referring to assessment scenarios mediated with a wider range of learning technologies and devices. In particular, our interest is focused on the use of different mobile devices such as smart-phones, PDA’s, consoles or supporting technologies such as NFC or RFID tags, GPS and Bluetooth to support more authentic assessment (Santos et al., 2013, de-Marcos et al., 2010; Hwang & Chang, 2011; Susono & Shimomura, 2006; among others). The use of appropriate technologies increases the possibilities of creating new types of questions, tests and assessment activities based on tests more adapted to real life interactions, situations and locations. The word computing covers the different ways of technologically representing and manipulating a test. See Figure 1.

![Figure 1. Relations between the assessment research contexts in TEL](image)

Mobile devices include a variety of technology, apps and features: wearable devices, tablets, smartphones, GPS, smart watches, google glasses, camera, pedometer, etc. The use of mobile technologies to enhance learning has been identified as beneficial for encouraging the development of meta-level thinking skills (Facer et al., 2004; Wegerif, 2002). These devices allow the user to interact with physical objects, physical locations, digital information, and with other users. We found really significant the statement made by Sharples et al., (2005): “Context is constructed by learners through interaction: To explore the complexity of mobile learning it is necessary to understand the contexts
Considering this reflection, we agree with Sharples that mobile devices can be adequately employed to allow users to interact with the physical/virtual context in a more realistic way. But mobile technologies are not enough, as we have discussed in the introduction; an adequate pedagogical plan is needed in order to guarantee the success of the assessment activity. We state that there are two important issues have to be considered when designing the pedagogical plan: First of all, the learning/assessment objectives have to be aligned with the performance of meaningful interactions. This means that the plan of the activity itself has to consider the kind of actions (interactions) done by the learner in order to solve the problem. Second, monitoring all the actions done by the learner can be difficult, but the higher the monitoring is the higher support can be provided to scaffold the learner during (and after) to understand which meaningful interactions has to perform in order to solve the activity.

Supporting meaningful interactions through scaffolding mechanisms

In this paper we want to place an emphasis on the importance of supporting “meaningful interactions”, this means to understand the potential of mobile devices to help learners to interact adequately with the environment in order to support more realistic tasks. As Dourish (2004) states, computational systems can be made sensitive and responsive to the setting in which they are harnessed and used. Mobile and sensor technologies play an important role monitoring interactions in context-aware activities. We agree with Dourish’s approach that the context can be seen as an interactional problem. Specifically, we analyse the importance of the context as interactional problem when designing an m-assessment activity.

As Woo & Reeves (2007) state: “Interaction in learning is a necessary and fundamental process for knowledge acquisition and the development of both cognitive and physical skills.” But these authors also identify in their paper the notion that the use of authentic tasks is not a guarantee of the adequate performance of meaningful interactions. One of the factors that help to guarantee the success of the activity is the application of scaffolding strategies. Hillman et al. (1994) insisted that all interaction in TEL is mediated via a medium. For this reason, it is necessary to understand the potential of mobile technologies (as a medium) to improve the interaction possibilities for assessment purposes. Engeström (based on Vygotsky’s idea of mediation) also discusses the importance of mediation as the use of artifacts, and/or concepts used by an actor, and how tools and the social context influence the actor-structure interactions (Engeström, 1999). Furthermore, as Cook (2010) states, mobile devices can be used as mediators in an ACD using them as the more capable peer can guide and scaffold the learner to find the adequate solutions. As Muirhead and Juwah (2004) identified, interaction is a dialogue between two or more participants and objects (i.e., physical or virtual) which occurs synchronously (or not) mediated by response or feedback and interfaced by technology. If we analyse the sentence, it can be applied to a process of answering questions in a test, where learners have to interact with information (e.g., physical, virtual or both) in order to provide an answer, and s/he receive feedback of this action by obtaining a response or a grade. Some researchers (e.g., Chen et al., 2003; Hwang and Chang 2011; Melero et al., in press; described with more detail below) have demonstrated how capturing information about location, dialogues among peers, time, position, interaction with certain objects, context, etc; can be used to analyse in real time the actions done by the learner and scaffold her steps during the activity.

We don’t forget that in an authentic real-world situation it would be unthinkable to have scaffolding assistance, however in this case we are talking about formative assessment situations where it is important to support the acquisition of the necessary learner’s skills.

Below we discuss which elements have to be considered to allow the learners to increase their interaction possibilities within a real learning context, and how by providing certain scaffolding strategies the user can be guided to engage with certain authentic tasks (by interacting with real objects, digital informal and other peers) and obtain specific results that can be monitored and assessed. In the following section we propose a model, called m-AssIST, that captures the necessary emergent properties to design an m-assessment context.
m-AssIST: Mobile - Interaction Scaffolding Temporal model

We claim that the use of mobile devices allows the learner to provide more authentic ways of answering questions by undertaking specific actions/interactions (not the traditional selection of a choice in a multiple choice question) which are associated with specific environments (i.e., locations and objects) and people (i.e., peers or people related with the context). These interactions with specific physical objects in an environment (i.e., measure the different angles of a room, run a distance in a given time, find and analyse a specific plant in a botanical garden, etc.) can be automatically monitored and assessed by the m-assessment system. The monitored actions can be used in favour of formative assessment by supporting the learner with scaffolding strategies during and after the assessment activity. However, in order to provide adequate scaffolding strategies for supporting the performance of meaningful interactions it is necessary to design adequately the pedagogical plan of the activity before its execution. We state that by using assistance and scaffolding mechanisms the dialogue between the learner and the m-assessment system can be mediated but some emergent properties have to be considered in order to guarantee the success of the activity.

In Figure 2 we present the mobile-Assessment Interaction Scaffolding Temporal Model (m-AssIST), this is an extension of the Context Model proposed by Tarasewich (2003). M-AssIST considers the different interaction possibilities during the temporal axis of an assessment activity (i.e., before, during and after the activity): (a) interactions with other learners, (b) interactions with the environment (i.e., indoors/outdoors spaces), (c) interactions with physical objects encountered in an environment, (d) interactions with the assessment activity, and finally (e) interactions with augmented digital information shown to support the activity.

As we illustrate in the model, the performance of an assessment activity has a temporal axis, this means that before, during and after the activity the m-assessment system has to use the learning objectives of the activity and the outcomes from the monitoring process and scaffold the learner to ensure the performance of meaningful interactions. The m-AssIST model has the potential to be useful especially when practitioners and technicians are in the phase of designing an m-assessment activity (this means, Before the Activity). It is at this moment that the pedagogical plan of the activity has to be completed. From our understanding the pedagogical plan has to consider the
learning/assessment objectives of the activity (i.e., HOTS to be assessed). Gulikers (2004) in his model reflects the need of having an alignment between the authentic learning tasks, the context (physical and social) and the learning goals (skills to be assessed). We recommend using Bloom’s Taxonomy (Bloom & Krathwohl, 1956) to indicate the type of HOTS that practitioners can assess with respect to a specific learning objective. Level 1: Knowledge; Level 2: Comprehension; Level 3: Application; Level 4: Analysis; Level 5: Synthesis; and Level 6: Evaluation. The higher a level is the most sophisticated, and this is the task (set of interactions) that the user has to do in order to answer the question. As we have already discussed in the previous sections, the feature of mobile devices allows the learner to be more active and this can be used in benefit of the assessment process. However, it depends on the practitioner to take profit of this benefit for assessing the learners’ HOTS. In relation to these approaches, our model makes emphasis on the importance of supporting the learners’ interaction with the context of the activity in order to provide the adequate assistance to scaffold her learning.

In this paper we claim that the pedagogical plan has also to include an analysis of the emergent properties (see Figure 2) that have an effect over the context of the designed activity. An important question to be answered is: which meaningful interactions the learner has to perform within the environment selected (i.e., the space where the activity is performed, for instance a Botanical Garden), with specific physical objects (i.e., plants, tools, people, etc.), with other learners (if collaboration is necessary) and with the augmented digital information provided by the system in order to solve the problem adequately?

By observing the m-AssIST model we see that the Interaction factor is in the middle of the figure. In this sense, for us the Interaction context is seen as the influential physical and social context discussed by Vygotsky (1978) or the relation between the physical environment and the interpersonal interactions proposed by Cook (2010). The adequate interaction of the learner with all the different factors considered in the context of the activity is essential to achieve the learning/assessment objectives of the activity. When the necessary interactions are identified by the practitioner these can be considered by the system in order to provide scaffolding assistance during the activity if some interaction is missed. For this reason a bold arrow goes from the Activity to the Learner, showing that a scaffolding assistance have to be provided to the learner if necessary. The missing interactions can be identified by the mobile system by monitoring the actions done by the student in real time or by considering the incorrect answers. In this case, the m-assessment system will be acting as the more capable peer.

m-AssIST is especially useful during the design phase of a m-assessment activity, but the model can also be used to analyse existing scenarios. In the following section we analyse existing m-assessment scenarios which accomplish the following characteristics: they are framed in a realistic context to assess higher order skills related to a subject matter, learners use mobile devices to interact with physical objects, digital information and peers, the students’ interactions are monitored and assessed in real time, and finally some guidance or scaffolding mechanism is provided to support the student. Specifically, we apply our m-AssIST model with the goal of analysing the examples selected identifying the methods (or limitations) used to guide and scaffold the learner in order to support meaningful interactions with assessment purposes. The corresponding analysis can be used by other practitioners to replicate strategies of an existing scenario, or to improve the identified limitations.

Authentic tasks, meaningful interaction and scaffolding in m-assessment

Three illustrative real examples have been selected in order to validate the fitness-for-purpose of the m-AssIST model from the point of view of using the properties to reflect about the necessary meaningful interactions in the authentic context and the adequateness of the scaffolding mechanisms used to provide assistance before, during and/or after the activity. The first two examples have been selected by using the combination of tags ‘Mobile, Assessment, Scaffolding, Authentic’ in Google scholar, then we selected two papers by taking into consideration the significance of the papers in terms of their scholarly impact (higher number of citations), different authors and designers, sufficiently different authentic environments (a natural park and an oldest part of a city), subject matters (natural science and local culture) and different skills assessed (further detail below). The third example also accomplishes the main characteristics of the first examples, but has been selected because some of the authors of this paper were involved in the design, creation and analysis of the m-assessment system and scenarios described and, therefore, are especially familiar with its definition and details. We use the emergent properties captured in the m-AssIST model to analyse the benefits and limitations of the selected examples in terms of: authentic tasks,
meaningful interaction and scaffolding mechanisms. We claim that if we show the value/utility and “usability” of the model for three significant and sufficiently different scenarios we can also discuss its transferability to other cases.

Analysis of existing examples

The first example is illustrated through the Bird-Watching Learning (BWL) (Chen et al., 2003). BWL was designed to assess students of natural science in preliminary school level when they are in an outdoors environment observing birds (in this case birds acts as physical objects and the natural park is the environment). PDAs and a Wi-Fi based wireless network are used as mediator technology during the activity. The assessment activity consists of the following tasks: while the students are observing the birds, a specific bird is selected by the instructor and students have to find the real specie selected by themselves. So, the learning objectives of the activity include assessing the following HOTS: analysis in situ of the information provided by the system, taking information apart (inspect, examine) from the exploration of the environment, and compare (differentiate, identify, resolve) this information in order to find the correct birds. As shown in Figure 3, HOTS (in form of verbs) are associated to the Activity property, this classification can be used to understand how to relate the activity with the other properties of the model. The BWL system has a testing subsystem that evaluates each learner’s level of learning/skills and attitudes. The instructor delivers information from the birds to the students in real time, so all the students have the same chance to see the related information (introduction texts are used as scaffolding mechanism before stating the main activity). The instructor can select 10 birds’ questions from the records, and transmit them to the students. By taking into consideration the responses, the system provides ‘different levels of assistance’ (manipulated by the instructor as scaffolding during the activity). The system provides ‘static images and hints’, and the students use this information to analyse, and explore the environment and discover the birds. The scaffolding level is determined according to the level of ability (application of HOTS) of each learner. The authors of the BWL system apply the following scaffolding strategy: the task is decomposed into (1) hierarchical tasks according to the skills. According to the step 1, the instructor (2) classifies the amount of support in decreasing levels and sets up a (2) repetitive authentic practice. The 3 steps are evaluated through an (4) ongoing assessment process (i.e., test questions are delivered to assess the accomplishment of tasks). This case is a very good example that shows the benefits of using scaffolding mechanisms to support meaningful interaction with the environment and physical objects (birds) whilst making possible a formative assessment activity. However, the application of the m-AssIST model allows us to detect some limitations in this scenario (see Figure 3).

Due to the technological limitations in terms of mobile devices in 2003, some mechanisms to monitor the activity in real time and support automatic assessment and scaffolding during the activity could be improved nowadays. For instance, instead of textual information, augmented simulations of the birds could be shown in the mobile device to help learners to achieve the HOTS, for instance identifying the correct bird. As used in Savannah (Facer, 2004) audio simulation of birds singing can be used to guide the students to find the adequate positions where they have to find the species. The interaction among learners is only referred to pairs of learners; however data from other peers could be used to improve the scaffolding assistance. For instance, the information from successful peers (pairs of learners who answered correctly the questions) could be used to scaffold other unsuccessful peers by sharing the location of those learners where they have found the correct birds. Finally, it would be useful to save the interactions undertaken by the students during the activity in a log and provide statistics after the activity to help the learners to reflect in their performance.

In the second example selected, Hwang and Chang (2011) propose a Formative Assessment-based Learning guiding Mechanism (FAML) to learn in situ about local culture (the Chin-An temple in southern Taiwan). In this scenario, fifth grade students used PDAs in a wireless network outdoors and indoors environment, they had to interact with the temple environment in order to answer related questions. In this case, the required HOTS were: students have to apply the theory learnt at classroom, classify the provided information and interpret this in a new situation (i.e., the real location) in order to solve the different questions (see Figure 4). In this case the system plays the role of the “more capable peer” guiding the students during the activity. In terms of a scaffolding strategy, the system provides “hints” and other ‘extra multimedia content’ to the students instead of giving an assessment of the validity of answers. One of the main benefits of the FAML system is that it was designed to encourage the student to interact with the environment instead of being focused on providing correct answers. When a question is answered incorrectly, instead of providing the correct solution the system encourages the student to make further interactions by providing more hints.
Using the emergent properties of the m-AssIST model (see Figure 4), we would like to highlight as limitation the fact of using the answers provided by the learners (correct or incorrect ones) as the unique mechanism to provide hints and guidance to the students. Nowadays, mobile devices allow the tracking of actions/interactions done by the user in spaces and with objects. As Zhou et al., (2008) show, tracking techniques can be used to detect changes in the viewer's position and reflect these changes in rendered augmented digital information. We envisage that calibration and registration tools could be used to align the real world (i.e., position of a statue in the temple) with the virtual one (i.e., activity and augmented digital information) when the user view is fixed. Other limitations of this example are observed regarding the scaffolding strategies, in particular before and after the activity where any information is provided to help the learner to understand her goal at the beginning of the activity, or to reflect about her performance at the end.

As the final example, we use the QuesTInSitu m-assessment system developed by (Santos et al., 2011, Melero et al., in press). We have selected a particular scenario to analyse the features of the tool to design m-assessment activities. QuesTInSitu allows learners to answer geolocated questions in situ (in an outdoors environment using GPS tracker, and in indoors environments using manual selection of questions over digital maps). When learners are in the correct position, a question appears and they have to apply their knowledge by interacting with the environment and physical objects in order to solve the question. For instance, one example of activity was done in a botanical garden where learners were asked to find specific plants in the garden (by using only the descriptions of the plants). Students put in practice the following HOTS by using tools to measure, calculate and analyse the plant (i.e., the average length of its leaves), by previously differentiate and identify the plant from others in the garden (see Figure 5). As in previous examples, the level of interaction complexity (and the application of HOTS) depends on the task asked by the question. In the first version of the tool (Santos et al., 2011) automatic feedback delivery was used after answering a question as a scaffolding mechanism during the activity. However, after performing several activities in real contexts with students there emerged the need for including scaffolding mechanisms before and after the activity too. For this reason, a new version of the system (Melero et al., in press) was developed. This version provides to the learner scaffolding mechanisms before, during and after the activity. The learner can select an ‘introduction guidance’ section before starting the activity (to understand better the context and the goals of the activity), during the activity hints can be manually selected by the learner if his/her has difficulties finding the correct way for solving a question. In parallel, during the activity the system monitors and collects information related to the learner’s position, time and
answering attempts. This information is shown to the learner as graphics after the activity. The goal of this graphics is to help the student to understand and reflect about their performance during the activity, for instance: Where she had more interaction difficulties to solve a question, or where the student interacts adequately.

**Figure 4.** m-AssIST model applied to the FAML scenario

**Context:** Formative assessment to learn local culture using PDAs and WiFi

**Figure 5.** m-AssIST model applied to QuesTInSitu

**Context:** Assessment in situ using smartphones and QuesTInSitu
By applying the m-AssIST model, we have realized that the QuesTiInSitu system could be improved by including more facilities to support the interactions among learners. All the activities performed with this tool were carried out by groups of students (2 or 3 using the same smartphone) and other groups doing the same activity in parallel, for this reason it is important to consider the interactions among the members of the same group and with other groups. In addition, the interaction with indoors environments could be enriched by integrating indoors positioning sensors to the system. As Sichitiu, & Ramadurai (2004) show in their study, wireless sensors networks have the potential to become the pervasive sensing technology of the near-future. In particular with assessment purposes, sensors will allow us to collect specific interaction information in a space but also manipulating an object (i.e., plants inside a greenhouse of the botanical garden), this information will be especially useful to support specific HOTS.

Conclusions

This paper has discussed the benefits of using mobile devices to support authentic assessment. The paper highlights the importance of using scaffolding mechanisms to assist learners when they interact with the environment, physical/digital objects, other learners and the activity itself. The selected scaffolding mechanisms have to be applied to facilitate the performance of meaningful interactions which is the based to support authentic tasks and HOTS (Woo & Reeves, 2007).

In this context, the m-AssIST model is presented as a model that captures the emergent elements of an authentic m-assessment activity. The model highlights the importance of considering the different interactions possibilities that learners have during a test-based activity. Monitoring those interactions through the mobile device allows the provision of specific scaffolding mechanisms before, during and after the activity. We have shown how the inclusion of scaffolding strategies transforms the m-assessment system in a ‘more capable peer’ assisting the learner during the activity to enhance the performance of meaningful interactions and as a consequence the correct achievement of HOTS. In this sense, the identification of the specific HOTS related to the Activity have to be considered in order to understand which specific features of mobile devices are going to be used. For instance, in order to support the identification of a bird, a layer of augmented reality in the natural park can be used to guide the student; or the mobile phone can be used in a group of students to explore different positions of the botanical garden and debate in real time the correct location of a plant.

We have used the m-AssIST model to analyse the characteristics and limitations of three existing m-assessment activities designed to support meaningful interactions in real environment and with physical/virtual objects. We would like to clarify, on the one hand, that other m-based test cases (different subject matters or learning levels) could be also analysed through the m-AssIST model. The model can be used to analyse existing scenarios (and systems), but also to design new ones and help practitioners and technicians to reflect which are the scaffolding mechanisms that can be used to support the particular interactions of an m-assessment context. On the other hand, although the examples selected were focused on the use of PDAs and mobile phones, other mobile devices can be used to assess the learners’ authentic tasks in real settings. For instance, Clark et al., (2012) provide a good example that shows how Microsoft Kinect can be used for assessment of postural control, recording the actions done by the users in a space giving the opportunity of using the entire body (and not only the hands) to make interactions that are computationally assessed. This last works links with our future research interest which deals on analysing the usefulness of mobile technology to support affective computing and intelligent interaction with assessment purposes.

Acknowledgements

This work has been partially funded by the Learning Layers project. Learning Layers is a 7th Framework Large-scale integrating project co-funded by the European Commission and the EEE (TIN2011-28308-C03-03) project.

References


A Web-Based Course Assessment Tool with Direct Mapping to Student Outcomes

Walid Ibrahim¹, Yacine Atif*, Khaled Shuaib¹ and Demetrios Sampson²

¹College of Information Technology, United Arab Emirates University, UAE // ²Department of Digital Systems, University of Piraeus, Greece // walidibr@uaeu.ac.ae // Yacine.atif@uaeu.ac.ae // k.shuaib@uaeu.ac.ae // sampson@iti.gr

*Corresponding author

ABSTRACT

The assessment of curriculum outcomes is an essential element for continuous academic improvement. However, the collection, aggregation and analysis of assessment data are notoriously complex and time-consuming processes. At the same time, only few developments of supporting electronic processes and tools for continuous academic program assessment and curriculum performance feedback have emerged. In this paper, we introduce a novel course assessment process supported by a Web based interface that articulates and streamlines the assessment data collection, performance evaluation and tracking of remedial recommendations. To close the assessment loop, the Web interface provides also a mechanism to follow up on the implementation of remedial recommendations and analyzes their associated reflective actions during the subsequent course assessment cycle. A guide to map assessment instruments to the course and overall program outcomes is advocated by the proposed tool to propagate the course assessment results towards higher educational objectives (e.g., student outcomes) in a dashboard-like assessment interface. This approach streamlines improvements in education through reflecting the achievement of course outcomes on the achievement of higher educational objectives. In addition, the tool maps the course outcomes to the corresponding course outlines to facilitate the detection of areas where revisions in the instruction and content is needed, and to best respond to recommendations and remedial actions. We provide a methodical approach as well as a Web-based automation of the assessment process, which we evaluate in the context of our regular academic assessment cycles that have eventually led to a successful international accreditation experience. The collected assessment data shows a significant improvement in the achievement rate of the student outcomes after deploying the tool.

Keywords

Web-based services, Course assessment, Student outcomes, Accreditation

Introduction

Education assessment refers to all activities which pro-vide information to be used as feedback to revise and improve instruction and learning activities (Black & William, 1998). Assessment outlines the foundation for continuous quality improvement and is considered to be a key element of educational processes. Two educational assessment models can be identified in the literature, namely the curriculum-based and the outcomes-based models. The curriculum-based model assesses learning retention based on the quality of the curriculum presented to the students and the learning methods implemented by the institution. The outcomes-based model, on the other hand, focuses on what the students should know and can actually do after completing their study requirements.

Lately, a new trend in educational assessment has been observed as more academic institutions are moving away from the traditional curriculum-based assessment models towards outcomes-based ones (Gardiner, 2002; Harden, 2002). Few factors have contributed to this new trend including some recent studies, which show that even students enrolled in respected academic institutions often fail to exhibit fundamental understanding of basic concepts and fairly easy physical systems (Harden, 2007). This is mainly because curriculum-based models do not usually make clear statements as to what students are expected to achieve upon completing a program of study. Having a carefully designed curriculum and a highly qualified faculty do not necessary mean that students comprehend the offered material. Therefore, to improve the efficiency of learning processes, academic institutions are increasingly adopting outcomes-based approaches for curriculum design (Harden, 2002). This shift makes them focus more on assessing the expected outcomes of the educational experience rather than the quality of the offered curriculum.

To ensure continuous improvements, the outcomes-based model relies on routinely and objectively assessing some standard learning outcomes along a comprehensive assessment cycle. The main components of this assessment cycle are illustrated in figure 1. During each cycle, suitable assessment instruments (or tools) are used to collect relevant assessment data for targeted outcomes meant by a given assessment exercise. The collected data are then analyzed.
and compared against the intended objectives. An assessment cycle is closed once proper recommendations are elaborated and remedial actions are implemented to address revealed deficiencies.

This widespread interest in the outcomes-led model is also driven by the accreditation prospects of academic institutions. Wergin (2005) mentioned that accreditation was the topic of more than 1,300 journal articles between 2002 and 2004. He also argued that the recent interest in academic accreditation is mainly because it is the only organized way by which an institution can convey quality assurance to the public. The accreditation board for engineering and technology (ABET) also believes that obtaining accreditation makes an institution in IT and engineering disciplines a better choice for students because it indicates that the students will gain standard knowledge and skills necessary to be productive members in contemporary professional careers. ABET accreditation exercise puts a strong emphasis on the outcomes-based model. Recently, the Canadian engineering accreditation board (CEAB) has also updated its accreditation criteria to adopt the outcomes-based model as well (Brennan & Hugo, 2010).

Figure 1. Information technology assessment plan

Recently, Al-Yahya and Abdel-Halim (2013), discussed their Electrical Engineering Department successful experience with outcomes based assessment of their program, which led to ABET accreditation. They describe preparations and the procedures followed to implement an evaluation system for the program by incorporating assessment of outcomes as well as a continuous improvement mechanism to develop and enhance the program. Their assessment model is based on two processes, an outer loop to evaluate the Program Educational Objectives (PEOs) once every four years and an inner loop, more frequently executed to evaluate the course and program outcomes, typically every year (Rogers, 2012). A unified approach to support outcomes based assessment has been investigated.

In Europe, outcome based assessment also gained momentum in the recent years (Crespo et al., 2010). Crespo et al., (2010) provided information on the different initiatives taken to support a unified conceptual map for outcome based assessment. They proposed a theoretical framework to integrate several concepts including learning outcomes, assessment processes and learning units. This increasing adoption of the outcomes-based assessment model has subsequently resulted in a considerable debate on how academic institutions can best define, assess, and evaluate learning outcomes. In this paper, we propose to streamline these processes through systematic electronic services featured by a Web-based assessment tool that is currently in use to capture the course-outcomes assessment results of an Information Technology curriculum. This development revealed that the performance of this curriculum could be tuned effectively by closing the assessment loop shown in figure 1, through automatically tracking recommendations and their associated remedial actions.

The Web based tool has the following benefits: (1) integrates planning and assessment in a systematic way; (2) maintain assessment records; (3) keep assessment agendas on track, with scheduled tasks for different levels; (4)
apply a uniform assessment model throughout the educational enterprise; and (4) root-out curricular deficiencies in a systematic way. Throughout this paper, we integrated the following contributions: (1) a combined assessment model to streamline the intricacies involved across section offerings of a course within a large academic institution, (2) a computational formulation of assessment results collected at course levels and reflected at program levels, (3) a data model to arrange relevant assessment indicators into a composite structure, and (4) a portal application to automate assessment processes for stakeholders involved in assessment cycles. We also report a successful experience of integrating the proposed model and deploying the suggested portal application into an actual academic assessment exercise, which we evaluate through a case study. Adding to the above benefits, this work has a twofold research value. First, we consolidate several modules of the assessment process, which used to be addressed separately in the literature. In doing so, we introduce new parameters to regulate the assessment standards of an institution and integrate the associated modules into a comprehensive assessment system. Secondly, we observe and report the results of actual experiments to gauge the effects of this streamlined approach. The outcomes of these contributions support further solutions to expand the scope of technologically enhanced education processes, at large.

The rest of this paper is organized as follows. Section 2 motivates the current needs for an effective assessment management system. Section 3 reveals our course assessment model and related processes. Section 4 presents the organization of assessment data. Section 5 provides the architecture of the supporting tool. Finally, Section 6 concludes the paper with an evaluation of the proposed tool in the context of actual case studies and reports some of the related lessons learnt from this experience.

**The need for streamlining academic assessment**

While outcomes assessment offers great promises for improving student learning, existing processes for integrating learning outcomes, collecting resulting data and analyzing student performance are limited. They are typically labor intensive, paper-based, and often exclusively driven by accreditation visits or other ad-hoc considerations. For example, the University of Alberta, Edmonton, Canada, estimated that the preparation for the last CEAB accreditation review of their nine engineering programs had cost them over $1 M and required the collection of more than a ton of documents, and 16,000 man/hours of preparation time (Dew et al., 2011).

The above example demonstrates the massive amount of administrative load that needs to be carried out by the faculty of an academic institution, while preparing for an accreditation visit. Overwhelming the faculty with administrative work, especially at the early stages of enforcing assessment processes, might trigger resistance against any resulting recommendation for changes to existing course contents and teaching practices. However, securing the support and the commitment of the faculty is essential for a successful implementation of assessment processes. Therefore, to effectively engage the faculty, course assessment processes should be as simple and efficient as possible, and integrate seamlessly with curricular components and teaching practices. A streamlined approach to automate academic assessment processes, alleviate the induced intricacies and reduce the associated load.

Although closing the assessment loop is essential for continuous quality improvement of a curriculum, it is usually the most overlooked part, and is typically where the assessment efforts get disrupted at the various levels. For example, at the course level, our own experience indicates that course coordinators usually terminate the assessment process after reporting the collected assessment results. They do not usually close the assessment loop by analyzing collected assessment data and creating appropriate recommendations and remedial actions to address any detected shortcomings in the learning process.

At the program level, however, the main problem is usually related to compiling the enormous amount of heterogeneous data collected from different course assessment exercises and using different assessment instruments (e.g., survey results, exit interviews, etc.). This federation of assessment data across program courses needs to be further converted into useful information that accurately reflects the achievement levels of student outcomes to facilitate curricular decisions. Failing to do so, may affect the correctness of the assessment results and could lead to ineffective or even wrong remedial actions.

Therefore, there is currently a clear need for automating the assessment workflows in higher education institutions. An assessment management system could effectively streamline the collection and analysis of assessment data. It will also contribute to lowering the complexity of the assessment processes, reducing the administration load.
assigned to faculty, and seamlessly weaving assessment exercises into existing teaching methods. Unfortunately, there are only few developments of such electronic processes and tools for continuous program assessment and feedback (Dew et al., 2011; Kerr, 2011; Essa et al., 2010; Booth, 2006). However, most of these systems have been designed in an ad-hoc way fulfilling the needs of a specific institution. For example, an accreditation management system has been developed and customized for the college of engineering at the University of Alberta, to satisfy the CEAB’s accreditation requirements (Dew et al., 2011). Similarly, an ABET course assessment tool (ACAT) system (Essa et al., 2010) was designed and developed to assist faculty at University of Nevada, Reno, in producing course assessment reports for ABET accreditation. Although it was designed for ABET accreditation, ACAT system deals only with the assessment at the course level and does not propagate the course assessment results to student outcomes (SOs) or PEOs.

Electronic assessment or e-assessment has been used in the evaluation of educational processes. E-assessment was defined in (Joint Information Systems Committee, 2007) as “the end-to-end electronic assessment processes where information and communication technology is used for the presentation of assessment activity, and the recording of responses. This includes the end-to-end assessment process from the perspective of learners, tutors, learning establishments, awarding bodies and regulators, and the general public”. Although e-assessment is mostly associated with assessment at the course level, the idea can be propagated to program assessment, in which a course assessment plays a cornerstone role. One of the main advantages of the use of e-assessment is attributed to the automatic and instant feedback, which is generated based on collected and processed data, to prompt parties involved in the assessment cycle to take necessary actions based on their predefined roles. Thus, the idea of using computerized tools and processes to effectively automate the assessment cycle of a particular program is appealing as it prevents information overload, limits roles of stakeholders to essential tasks, instills process uniformity and prompts just-in-time actions.

**Figure 2. Information technology program assessment hierarchy**

**Assessment methodology**

We have recently implemented a comprehensive assessment and evaluation system as part of our efforts in preparing our Bachelor of Science in Information Technology program for accreditation by ABET. However, this effort resulted in a sustained activity beyond the accreditation goal. The assessment system consists of three main processes based on ABET guidelines (Sanderson, 2009) as shown in figure 2. At the lower level, the course assessment process is used to measure the achievement of course outcomes (COs). COs describe the knowledge, skills, and/or competencies that the students should have or be able to demonstrate upon completion of the course (Yue 2007). COs assessment results are then combined with other program-level assessment tools (e.g., student survey, exit exams, exit interview, etc.) to measure the achievement of the SOs at the next stage. SOs describe what students are expected to learn and be able to do by the time of their graduation (Sanderson, 2009), which include cognitive, affective, behavioral, social, and ethical performances (Wyne, 2010). At the third level, the PEOs (Williams, 2010)
are measured and evaluated. PEOs are broad statements that describe the strategic career and long-term professional accomplishments the program is preparing its graduates to achieve 3-5 years after graduation.

Although COs were regularly assessed by the end of each course-offering since the deployment of our curriculum in 2005, the legacy course assessment process had several drawbacks that diminished its effectiveness, which include:

- Missing a formal evaluation and feedback mechanism.
- Missing a standard mechanism for compiling the assessment data collected from multiple section offerings of the same course. The assessment of COs was done at the section level, where instructors assess the outcomes based on their individual section assignment, and submit their assessment reports to the course coordinator.
- The assessment data was subject to class gender (as separate male and female classes are scheduled in our institution), and the number of students per section.
- No unification of assessment tools at the course level. Different assessment tools were used by different instructors to evaluate the same CO. Therefore, it was difficult to compile the assessment results together to generate the overall course assessment report.
- Lack of automation. Although softcopies of all assessment reports (PDF format) were regularly stored in a dedicated assessment repository, it was very difficult to extract the assessment results (automatically) and propagate them across a higher-level to assess strategic objectives (e.g., SOs, and PEOs) as shown in Fig 2. It was also difficult to track the corrective recommendations and their implementation to address remedial actions and link them to their assessment origins.

To overcome these drawbacks, a new assessment model has been developed by the assessment committee of our College of IT, and subsequently approved by the College council during the 2008-2009 academic year (see figure 3). The new assessment process starts by defining COs for a new course or revising existing ones based on ABET standards for an IT curriculum and Bloom’s taxonomy for cognitive skills. The curriculum committee at the College level is responsible for approving any alteration in COs requested by the corresponding course coordinator.

As part of the proposed model, a course assessment process requires that a course committee routinely convenes each time the course is offered. The course committee consists of the course coordinator (chair) and all other instructors who are teaching sections of that course during that offering. The course committee meets at least twice during the course offering. The first meeting is usually scheduled before the end of the first week of a course offering period. In
this meeting, the course committee (i) discusses the assessment results and recommendations from the last course offering, (ii) decides on a set of remedial actions to address the advocated recommendations in the previous assessment exercise, and (iii) agrees on the set of tools that will be used to assess each outcome (see figure 3 for examples of course assessment tools).

During the course offering, each instructor teaching the course is responsible for collating the assessment data for his/her own section(s) and preparing a simplified section assessment report. The section assessment reports are then sent to the course coordinator within two weeks after the administering the final exam. To minimize the assessment effort, the section assessment report includes only the following information:

- The number of students
- The assessment tools used
- The mean and standard deviation for each outcome achievement level
- Any assessment related remark (optional).

Although bi-modal distribution could be more appropriate in some situations to model the students’ performance, the tool assumed a Normal distribution for the students’ performance to facilitate the aggregation of the assessment data from different sections and different assessment instruments producing the overall CO assessment results. To calculate the mean and standard deviation for each outcome, the course coordinator aggregates the performance of the students in each used assessment tool. We assume that each course \( c \) has a set of outcomes \( O_c \), a set of assessment tools \( T_c \), and is offered to a set of sections \( S_c \). Therefore, the mean and standard deviation for outcome \( o \) is calculated as follows:

\[
\mu_{soc} = \frac{\sum_{t \in T_c} \mu_{toc} \times \alpha_{to}}{\sum_{t \in T_c} \alpha_{to}} \quad (1)
\]

\[
\sigma_{soc} = \sqrt{\frac{\sum_{t \in T_c} \sigma_{toc}^2 \times \alpha_{to}}{\sum_{t \in T_c} \alpha_{to}}} \quad (2)
\]

Where \( \mu_{soc} \) and \( \sigma_{soc} \) are the mean and standard deviation when tool \( t \) is used in section \( s \) to assess outcome \( o \) of course \( c \), \( \alpha_{ts} \) is a mapping factor that determines the contribution of the assessment tool \( t \) to the achievement of outcome \( o \) such that \( \sum_{t \in T_c} \alpha_{to} \leq 1 \).

The course coordinator compiles the received section assessment reports, and calculates the aggregated course level assessment results. Using the mean and standard deviation for measuring the achievement of the outcomes at the section level facilitates the aggregation of the results from different sections to calculate the overall course assessment, regardless of the assessment tool used in each section. For each outcome \( o \) of course \( c \), the aggregated mean and standard deviation are calculated as follows:

\[
\mu_{oc} = \frac{\sum_{s \in S_c} \mu_{soc} \times n_s}{\sum_{s \in S_c} n_s} , \quad (3)
\]

\[
\sigma_{oc} = \sqrt{\frac{\sum_{s \in S_c} \sigma_{soc}^2 \times n_s}{\sum_{s \in S_c} n_s}} , \quad (4)
\]

where \( n_s \) is the number of students in section \( s \). Assuming Normal distribution, the achievement level of outcome \( o \) of course \( c \) (noted \( A_{oc} \) below) is calculated as the percentage of students who scored above a predefined cutoff threshold \( \lambda \), as shown next:
\[ A_{\text{co}} = 0.5 - 0.5 \times \text{erf} \left( \frac{\lambda - \mu_{\text{co}}}{\sigma_{\text{co}} \times \sqrt{2}} \right) \] \quad (5)

For example, assume that \( \mu_{\text{co}} = 0.74 \), \( \sigma_{\text{co}} = 0.093 \), and \( \lambda = 0.7 \), the outcome achievement level in this case is the percentage of students whose score is above \( \lambda \), which in this case is 66.64%.

To close the assessment cycle, after aggregating the section assessment results, the course coordinator convenes the course committee for the second meeting. During that meeting, the committee reviews the remedial actions implemented during the last offering and evaluates their impact. The committee also compares the calculated achievement level for each outcome (\( A_{\text{co}} \)) against the targeted standard set by the College administration. Both the targeted achievement level and the cutoff threshold (\( \lambda \)) are proposed by the curriculum committee and approved by the College council. If any deficiency is identified, a new set of recommendations is created. After the second meeting, the course coordinator uses the online course assessment tool, presented later in this paper, to fill and submit the course assessment report.

It is important to mention that the effectiveness of the assessment process is very sensitive to the cutoff threshold (\( \lambda \)) and the achievement target level. These two values should be carefully set as they affect the continuous quality improvement cycle. For instance, setting either of these two values inappropriately low implies that there is a high probability that the outcome achievement level will be satisfied. Consequently, no recommendation or remedial actions will be needed, which stops the continuous improvement cycle prematurely. Therefore, if the CO targeted performance level is satisfied in two subsequent assessment cycles, the course coordinator might consider raising the performance bar. This is done by submitting a request to the curriculum committee for review, recommendation and approval. The curriculum committee may set different values for different courses depending on the course type and level. For example a targeted performance level for a mandatory core course in the curriculum could be set higher than the value used for an elective specialized course.

**Course and assessment databases**

To organize and control the access to course related information, a course database has been developed. The database is used to store all course related attributes such as catalogue description, textbook, credit hours, topical outline, grading criteria, pre-requisites, outcomes, and the name of the course coordinator. Access to this database is geared by a portal-based application, which sets proper authentication across different course-related stakeholders (faculty vs. program coordinator). The course database also maintains a mapping between the outcomes and the course topical outline. This mapping helps the course coordinators analyze the collected assessment data as it highlights which part of the course contributes to which outcome. As a result, course coordinators are able to make more relevant and effective recommendations by identifying the areas that need more attention when a deficiency in one of the outcomes is observed. The effectiveness of the recommendation is linked to the impact of the implemented remedial actions as observed by the instructors.

To simplify the collection and extraction of COs’ assessment results, an assessment database was created. The assessment database is also used as an assessment data warehouse to keep a history of COs assessment related data across several course offerings. For each assessment cycle, this database stores the following information:

- total number of students,
- applied assessment tools,
- aggregated mean and standard deviation for each CO,
- assessment remarks for each CO, which might include a comparison between the performance of the students during the current offering and the previous one, or a comment on the effect of an remedial action.
- new recommendations,
- a description, status, and impact for each implemented remedial action,
- mapping between each new recommendation and the related COs, and
- mapping between the remedial actions and the previous recommendations.
The mappings in items 7 and 8 are paramount as they effectively document the closure of the assessment loop. For each deficiency, the database is used to retrieve the assigned recommendations and the remedial actions taken during the following assessment cycle. It can also be used to track the implementation of each remedial action and their impact. This allows course coordinators to adopt or reject an implemented remedial action based on its observed impact on the CO achievement. Remedial actions with a negative impact would be replaced by different ones during the following course offerings. In addition to the mean and STD stored in the database and used later for the calculation of the SOs, the performance of the students in the different assessment instruments used in the calculation of the assessment results are collected and stored in a separate repository.

The assessment database also maintains a mapping between COs and SOs, which allows the tool to utilize the COs’ assessment results as an instrument for assessing SOs. Each CO may contribute to zero or multiple SOs, while one or multiple COs may contribute to the same SO. The CO contribution level to a specific SO depends on the number of course-related contact hours dedicated toward the achievement of the CO which is mapped to that SO. Assuming a 15 weeks course offering, a 3 credit-hours course can produce up to 45 contact hours to cover the different COs throughout the course offering. Therefore, assuming that the curriculum has a set of student outcomes $O_S$, and each course $c$ has a set of outcomes $O_C$, the mapping function $M_{ij}$ defines the number of contact hours a course outcome $i \in O_C$ contributes to a student outcome $j \in O_S$ such that:

$$\sum_{i \in O_C} \sum_{i \in O_S} M_{ij} \leq 45. \quad (6)$$

The course and the assessment databases form the back-end of our Web-based assessment tool. The tool is part of an integrated assessment portal that was designed and implemented during the preparation for the successful ABET accreditation exercise of our BS in IT program. The implementation details of this tool are further revealed next.

**Web-based assessment tool**

The Web-based assessment system is implemented on top of our existing curriculum management application developed as part of a College-wide portal. This allowed us to integrate the academic assessment operations with some existing services such as online syllabi and (automatic) study plan generation for individual students (as part of our student advising process). The portal is implemented using Liferay portal development platform (Sezov, 2011). The selection of Liferay is due to its open architecture, which could also be seamlessly integrated into Course Management Systems like Moodle. We preferred this approach to separate course assessment from content management in order to facilitate change management and to link assessments with our College-wide admin processes that are already hosted in the Liferay-based portal.
Figure 4 shows the implementation architecture following a common layered-structure, which includes: presentation layer, business logic layer, data access layer, and databases layer. In portal context, the presentation layer is fundamental since it expresses the required workflows across the entire layer hierarchy. The presentation layer comprises components used in the interactions between academic assessment related personnel and the customized user interfaces. The major role of the presentation layer is to display the information required at each assessment level (i.e., COs, SOs, and PEOs) as well as to translate user’s instructions into business logic layer operations. The data access layer contains the collected raw assessment data and their organization into the portal-provided assessment database.

Using Liferay portal development platform, we generate customized presentation interfaces from our course and assessment databases. The portal which receives client requests, includes a portlet container. A portlet is a Web component (Java-based), which basically listens to client requests and generates dynamic contents accordingly using pre-defined templates. A portlet may accommodate several windows into a portal page. Figure 5 shows the portal page for a course coordinator. The page has three windows that allow the course coordinator to edit the course information (figure 5), edit COs mapping to SOs (figure 6), or assess the related COs (figure 7).
Using the first window (see Fig 5) the course coordinator may update the course textbook, topical outlines, grading scheme, or the mapping between the COs and the topical outlines. Other information such as the course name, title, catalogue description, credit hours, and outcomes can only be modified by the curriculum committee. Although multiple sections of courses could be offered simultaneously each semester by different instructors, only the course coordinator is allowed to edit the course information or update the mapping between COs and SOs, and COs to the topical outlines. This ensures a uniform course offering with a common set of COs across the offered sections. However, the tool allows different courses to have different number of COs and topical outlines to accommodate the offered variety of courses at the different levels.

A course assessment form (see figure 6) is made available for instructors to enter relevant data for the current assessment cycle. According to the course assessment process, the course coordinator should complete the course assessment form once in each academic year. In addition to the outcomes means, standard deviations, and the used assessment instruments, the form also collects essential information regarding the assessment loop closure. This information includes the remedial actions taken during the semester to address the recommendations from the previous offering. For each taken action, the status and the associated recommendations are also specified. The form also lists the recommendations created by the course committee to address any new shortcomings. Each new recommendation is mapped to one or more CO. By clicking on the save button, the course coordinator signs and saves the assessment form in the assessment database. The assessment database simplifies the assessment loop closure significantly. It is now straightforward to query the database to list all the recommendations created to address a shortcoming in a specific year to trace its impact. The action items created to address those recommendations, as well as the status of each one can also be retrieved and traced.

The third window is used to map COs to the standard “14 SOs” defined by our IT program. The portlet assumes that each course has a maximum of 45 contact hours, which can be used to contribute to the SOs (as described earlier). To simplify the mapping process, instead of specifying the number of hours, the course coordinator may set the mapping level to either none, some, substantial, or significant, which is respectively translated by the portlet to 0, 1, 3, or 6 contact hours. The portlet allows the course coordinator to set and adjust the contribution level of each CO, however it prevents any further contribution if the accumulated contact hours reached the 45 hours limit. The saved mapping information is then used along with the COs assessment results (mean, and STD) to automate the calculation of SOs achievement results.
Discussion

Since its deployment in Spring 2008, the assessment portal witnessed increased access frequency by academic personnel in our institution as shown in figure 8 considering there are about 32 course coordinators (some of whom are looking after multiple courses). Table 1 shows the assessment statistics for Fall 2009. It shows that 85.7% of the offered courses have at least the mean and standard deviation of the COs submitted, while only 50% of course coordinators used the assessment remarks to comment on the new assessment results. The number of courses with at least one recommendation increased from 16.67% in Spring 2009 to 69.1% in Fall 2009. Out of the courses with submitted recommendations in Spring 2009 (16.67%), only 52.17% have implemented at least one action item during Fall 2009.

Table 1. Fall 2009 Course assessment statistics

| Courses with COs Mean and STD | 85.7% |
| Courses with assessment remarks | 50% |
| Courses with previous recommendations | 16.67% |
| Courses with at least one action item | 52.17% |
| Courses with at least one new recommendation | 69.1% |

The tool was introduced to the course coordinators during one of the regular faculty meeting followed by a workshop on the new course assessment process and on how to use the tool to fill and submit a course assessment report. During the initial deployment of the tool, the existing curriculum related information (e.g., course titles, catalogue descriptions, credit hours, textbook titles, etc.) was easily collected and uploaded to the course database by administrative personnel. Course coordinators were then asked to complete the following related tasks: (i) revising current COs, (ii) mapping the revised COs to SOs, (iii) mapping the revised COs to course outlines, and iv) use the provided portal interface to upload these information to the course database. Setting the completion of these tasks as one of the milestones for ABET accreditation and with the proper and timely guidance, course coordinators were able to complete these tasks appropriately within the allowed time.

Figure 9 shows the percentage of courses with the assessment form completed by the course coordinator for the last four years. It shows that the percentage of the course assessment increased significantly from 34.2% in 2008-2009 to 85.5% in 2009-2010. We believe that this significant increase is due to two main reasons. First, the tool was deployed late in the second semester of 2008-2009. Hence, the tool was used only for a subset of the courses offered in the second semester. Second, the increased assessment and accreditation activities within the college during 2009-2010 in preparation for the ABET accreditation review had improved the assessment awareness within our College. Figure 9 shows that the percentage of assessed courses improved back to 80% in 2011-2012 (following the successful accreditation outcome), after a slight decline in 2010-2011. This observation indicates the sustained assessment efforts facilitated by our streamlined electronic approach.
Over the last four years we have overseen the implementation of the course assessment process and the usage of the web-based tool. We noticed that the main difficulty facing course coordinators was the collection of the assessment data and the calculation of the COs assessment results. No complaint has been received from the course coordinators regarding the usability of the web-based tool. Therefore, to overcome this difficulty, the assessment committee has developed new macro-based course assessment templates that automate the calculation of the assessment data from individual sections and aggregate them to calculate the COs assessment results. The templates were deployed during the 2013/2014 academic year with positive feedback received from course coordinators on how these templates significantly simplified the course assessment process. We are planning to develop an online version of the templates and integrate them within the assessment portal during the 2014/2015 academic year.

The impact of the new course assessment and its Web-based tool is twofold. First, it helps improving the quality of the curriculum by improving the quality of individual courses. Using the course outlines to COs mapping feature helps course coordinators to create effective recommendations that target areas of deficiencies. The evidence of this statement is shown in figure 10, a total of 114 recommendations were submitted in 2009/2010 academic year to address the identified deficiencies in COs assessment. A total of 57 remedial actions were implemented during the 2010/2011 academic year to address some of these recommendations, which exceeds by large margins of previous manual processes.

In this study the curriculum improvement is assessed by the achievement level of SOs. Figure 11 presents the percentage of SOs which successfully exceeded the 70% threshold for three consecutive assessment cycles. The percentage is calculated using only COs assessment results, and the COs to SOs mapping provided by the course coordinators. It shows a clear positive improvement in the achievement of SOs since the first assessment cycle after the tool deployment in the 2009-2010 academic year. It reveals that throughout that same academic year, six out of the 14 outcomes were below the 60% achievement-threshold target. While none of the SOs were below the 60% achievement threshold in the following two assessment cycles.

However, as shown in figure 10, only 47% and 52% of the recommendations developed in an assessment cycle were implemented using action items during the following cycle. This is considered a serious deficiency since implementing the recommendations through remedial actions and measuring their impacts is essential for closing the assessment loop. This deficiency usually occurs when the course coordinator ignores calling the course committee to meet at the beginning of the semester to discuss the previous assessment report and decide on the appropriate action items to implement, as outlined in the proposed assessment model.

The suggested Web-based application has also a positive impact on the achievement level of SOs. The tool automatically assesses the SOs by using the submitted course assessment results, and the mapping between COs and SOs maintained by the assessment database. This has prompted the academic assessment coordinator to use the course assessment results, for the first time during the 2009-2010 academic year as a tool to assess SOs. We believe that using the course assessment results as one of the assessment instruments has enhanced the accuracy of SOs.
assessment process, which used to depend only on the student performance in administered exit exams and surveys. The impacts of the provided Web-based assessment tool motivated an increasingly larger population of academics to embrace an outcomes-based education and in extending the dynamics of curricular revisions. The observations drawn from a three-years processing of assessment results illustrate improvements in outcomes achievement that are attributed to the provided Web-based assessment tool.

Conclusions

In this paper, we introduced a new course assessment approach and a supporting Web-based application to streamline the overall processing of collected assessment data. The Web-based application provides a user-friendly interface to the course coordinator, the curriculum committee, and the academic assessment coordinator, for accessing the assessment results as well as tracking the related recommendations and the status of the remedial actions.

One of the main objectives of the accreditation review is to ensure that students are achieving the intended outcomes through a continuous quality improvement process. The proposed tool has proved to increase the stakeholders’ rate of contribution into the academic assessment process. Based on the observed results, a consistently high rate of input is collected from course coordinators to impact the curriculum quality. This is an indication of the level of trust in the provided Web based tool to intervene in the process of optimizing the performance of academic programs administration. The opportunity to streamline the outcomes assessment process directly from course coordinators, has greatly contributed to identify areas of deficiencies to close effectively the assessment loop. This procedural engagement translated into remedial actions and related documentation to address and track the discovered deficiencies and to reflect back on the recommended enhancements.

References


A Reconfigurable Simulation-Based Test System for Automatically Assessing Software Operating Skills

Jun-Ming Su* and Huan-Yu Lin

Department of Information and Learning Technology, National University of Tainan, Tainan, Taiwan 70005 // Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan 300 // juming.su@gmail.com // linhuanyu@gmail.com

*Corresponding author

ABSTRACT

In recent years, software operating skills, the ability in computer literacy to solve problems using specific software, has become much more important. A great deal of research has also proven that students’ software operating skills can be efficiently improved by practicing customized virtual and simulated examinations. However, constructing these simulation-based examinations is cost intensive and time-consuming. Thus, a reconfigurable simulation-based test system designed on a proposed Software Operation-Finite State Machine has been developed. Using this system, teachers can use a graphic authoring tool to construct and reconfigure simulation-based tests for designated software operation tasks. For the students, these simulation-based tests can be performed via the Web browser without any extra computer environment setting costs. After students finish the test, the proposed system can automatically assess their online operating paths to generate the assessment report for self-reflection. Moreover, in order to share and reuse the designed simulation-based tests, this paper suggests some extensions of IMS Question & Test Interoperability to support the proposed simulation-based operation test format. Finally, an experiment was conducted to evaluate 38 sixth grade elementary school students. The experimental results show that this system can meet both teachers’ and students’ educational requirements while significantly improving students’ learning performance of the software operating skills.

Keywords

Computer-based testing, Software operating skills, Simulation-based test, Automatic assessment, Computer literacy

Introduction

With the trend of the digital age, the skill of solving problems using computer applications, a part of computer literacy skills, has become more and more important (Grant et al., 2009). These skills are also important in companies (Breivik, 2005; Grant et al., 2009) and schools (Dunn, 2002; MSACHE, 2003; NCCS, 2009; UCF, 2015). According to the research of cognition (Anderson, 1981, 1983), computer skills consist of a declarative knowledge aspect (i.e., the knowledge of understanding the computers and software applications), and a procedural knowledge aspect (i.e., the skills of operating these software applications) (Koong et al., 2014). In traditional learning environments, teachers usually evaluate students’ software operating skills (SOSs) using laboratory examinations. Due to the high costs associated with manual evaluations and testing environments, conducting a software skills test via a laboratory examination approach is costly and time-consuming. Previous e-learning and e-testing studies (Hu et al., 2009; QTI, 2012) contributed many ideas and technologies for online learning and testing of declarative knowledge, but those approaches are inappropriate for evaluating procedural knowledge of SOS due to a lack of practice and test mechanisms in computer operational software (Kinnersley et al., 2001; Koong et al., 2014). Thus, previous research (Kinnersley et al., 2001) proposed a multimedia-based virtual examination for learning and testing SOS. Some large software companies, such as Microsoft (2010) and Cisco (2004), as well as certification examination providers (TQC, 2015; Katz, 2007; NCDesk, 2009) have constructed their own online test or certification examination systems for the SOS. However, students require a large amount of various operational tasks for practicing and testing, but these systems do not offer functions for teachers to design new software operational tasks. Moreover, these certification exam systems, providing only summative assessments (Harlen & James, 1997), are not able to capture students’ operating paths. Therefore, teachers cannot further analyze and diagnose the operational behaviors to understand their students’ misconceptions and learning problems.

Accordingly, this study proposes and implements a reconfigurable simulation-based test scheme, which can assist teachers in efficiently constructing or reconfiguring a test for a designated software operating task using a developed graphical authoring tool. To simulate the software operational process, a Software Operation-Finite State Machine (SO-FSM) based on the concept of the FSM, is thus proposed, where the states with the background images represent the operating steps to simulate operating status and the transitions represent all possible operating processes.
and actions that can be performed. The generated simulation-based tests can be conducted via Web browsers without any extra test environment setting costs; students can virtually operate the simulated software to fulfill a given task. After the operation of the tests, students’ software operating paths (SOPs), as the transitions in the SO-FSM, can, thus, be captured by the system to allow teachers to further diagnose students’ misconceptions of software operation. Moreover, in order to share and reuse the designed simulation-based tests based on the SO-FSM, this study also suggests and defines some extensions of IMS Question & Test Interoperability (QTI) Specification, version 2.1 (QTI, 2012). This standard format for the representation of assessment content and results will be used to support the interoperability function. Based upon the aforementioned ideas and scheme, a Simulation-based Software Operating Test (SbSOT) system has been developed for automatically assessing SOSs. This system is then used for sixth-grade elementary school Computer Science teachers to construct and conduct software operation tests. The experimental results show that this system is capable of significantly improving students’ SOSs. The satisfaction degrees of teachers and students with respect to the SbSOT system are high as well.

Related works

In recent years, educational researchers proposed many kinds of multimedia learning approaches to promote active learning and to improve students’ understanding of real world problems using interactions and simulations. For example, let us mention the following authors. Feng et al. (2005) and Yu et al. (2002) used a multi-player game to support competition for learning fractions. Galvão et al. (2000) used a 3D simulation, where students could adjust parameters of the manufacturing process and observe the results to attain manufacturing process control skills. Patomäki et al. (2004) designed a virtual learning environment and real tools to train visually impaired young children. Rodriguez et al. (2007) used a siege game to assist students in learning and testing Newton’s laws of motion in physics. The experimental results of Holzinger et al. (2009) also showed that introductory movies and simulations could improve learning efficacy in medical education. Several studies (Zenisky & Sireci, 2002; Scalise & Gifford, 2006) also applied multimedia technology to test students’ skills in real world problems, which were difficult to describe in text-based questions. Quellmalz and Pellegrino (2009) showed that simulation-based assessments, such as a food web simulator, could assess inquiry skills. Basu et al. (2007) and Cheng and Bischof (2007) described that the assessment with various multimedia content and interactions can test not only students’ knowledge, but also cognitive skills, including linguistic, logical-mathematical, spatial, bodily-kinesthetic, musical, interpersonal, and intrapersonal. The review of Parshall et al. (2000) introduced some operational examinations, which required students to perform multiple steps in an operating process to assess the operating skills for solving a real problem. Koong et al. (2014) developed a computer-assisted learning (CAL) system to help students learn operational software. The results indicated that students’ learning effectiveness of the procedural knowledge can be enhanced significantly through the operational test items developed by collecting frequent software operation error patterns.

Thus, multimedia tests seem to be suitable for practicing and testing SOSs by simulating the tested software’s behavior and interacting with students. The online test mechanism can also reduce the money and manpower required for setting up a test environment and for monitoring the test process. Among previous studies of online software operating tests, Kinnersley et al. (2001) constructed a flash-based software simulator to test students’ operating skills. However, constructing this kind of simulator was time-consuming for teachers (Koong et al., 2014). Microsoft (2010) and Cisco (2004) both developed certification examination systems, using software simulators to test students’ skills of software and hardware control. In North Carolina, the North Carolina Online Test of Computer Skills (NCCS, 2009) used the NCDesk (2009) system to provide a computer operating skills test for eighth grade students. Educational Testing Service (ETS, 2015) proposed an iSkills™ (2015) test to evaluate students’ information and communication technology literacy (Katz, 2007; Somerville, 2008). Students were asked to use the simulation software, such as a search engine, word processor, or e-mail tool, to complete the tasks. These online certification examinations and flash-based simulators were successful in assessing students’ SOSs through online environments to reduce the cost of performing a large-scale software skill tests. Unfortunately, in general, teachers could not afford the high costs related to constructing these kinds of simulation-based tests for students to practice and test their skills. Moreover, these current studies and applications do not seem to monitor and record students’ detailed operating steps for further analysis, so teachers lacked sufficient information to identify the students’ misconceptions and to further improve software skills.
Reconfigurable simulation-based test scheme

According to our observation, most of the operating behaviors of popular operational application software, such as MS Office® and Photoshop®, include sequences of operations among forms, menus, and text areas. Therefore, this kind of software operation, referred to as form-based software operations, can be easily simulated through corresponding images of software screenshots for each operation with customized functions. Students can perform the software operations, such as click, key-in, and drag-and-drop, in an online virtual/simulated environment, e.g., Web browsers, according to the SOS test requirements. Consequently, the students’ operations can be recorded and used to automatically evaluate and assess their skills. Thus, according to the aforementioned idea, a reconfigurable simulation-based test scheme based on the FSM theory (Kandar, 2012) is proposed, called SO-FSM, and further applied to MS Word software skills test to demonstrate the efficacy of SOS evaluation.

Software operation-finite state machine

Because software reactions are deterministic, a sequence of replaced software operating screenshots can be used to easily simulate an operating process of the form-based operational software. Therefore, according to the FSM theory (Kandar, 2012), this kind of image replacement with the corresponding operations can be purely modeled as the transitions between different states in an FSM. Thus, the SO-FSM = (S, ∑, s0, δ, F) is defined to model the form-based software operating process, where (1) S = {s0, s1, …, sn} is a set of states denoting the operating statuses. Each state si = (imgi, descsi, pi) contains screenshot image imgi, description descsi, and state performance tag pi, where pi ∈ {Internal, Passed, Failed, Misunderstood} describes the status of the students’ skills in the test, e.g., students operate within the process of performing the skill (internal), successfully perform the skill (Passed), unsuccessfully perform the skill (Failed), or misunderstand the skill (Misunderstood); (2) ∑ = {e1, e2, …, em} is a set of input events, where ei can be a specific operation, such as “click” and “double-click” within a specific action area, or the event “Failed” to denote the undefined events. As shown in Table 1, the events are categorized into eight Web browser event classes used in this study, where E = {EClick, ERight-Click, EDouble-Click, EDrag-and-Drop, EHot-Key, EKey-In, EChoose-Menu-Item, EFailed} is a set of Web browser event classes and ei ∈ E; (3) s0 ∈ S is the starting state; (4) δ is the state-transition function; (5) F ⊂ S is a set of final states.

<table>
<thead>
<tr>
<th>Web Browser Event Class</th>
<th>Information Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EClick</td>
<td>(x, y), width, height: as integer data type</td>
<td>Describes the action area of mouse’s left-click, containing the information of (x, y) coordinates, width, and height.</td>
</tr>
<tr>
<td>ERight-Click</td>
<td>(x, y), width, height: as integer data type</td>
<td>Describes the action area of mouse’s right-click and the same event information of EClick.</td>
</tr>
<tr>
<td>EDouble-Click</td>
<td>(x, y), width, height: as integer data type</td>
<td>Describes the action area of mouse’s double-click and the same event information of EClick.</td>
</tr>
<tr>
<td>EDrag-and-Drop</td>
<td>(from_x, from_y), from_width, from_height, (to_x, to_y), to_width, to_height: as integer data type</td>
<td>Contains the begin action area (from_) and end action area (to_) of a drag-and-drop event.</td>
</tr>
<tr>
<td>EHot-Key</td>
<td>key: as string data type</td>
<td>Describes a specific hot key and hot key combination, e.g., ctrl+a.</td>
</tr>
<tr>
<td>EKey-In</td>
<td>text: as string data type</td>
<td>Describes the key-in string while a user fills in a textbox.</td>
</tr>
<tr>
<td>EChoose-Menu-Item</td>
<td>text: as string data type</td>
<td>Describes the corresponding string of a choose-menu-item from a drop-down list.</td>
</tr>
<tr>
<td>EFailed</td>
<td>None</td>
<td>Denotes the inapplicable events.</td>
</tr>
</tbody>
</table>

According to the effects defined in Table 1, students’ operations are able to be categorized into seven operation classes, as shown in Table 2, where let O = {OClick, ORight-Click, ODDouble-Click, ODrag-and-Drop, OHot-Key, OKey-In, OChoose-Menu-Item} is a set of operation classes. In order to further assess the operating steps, the sequences of students’ operations in the SO-FSM are recorded as an SOP, where SOP = <b1, b2, …, bp> is a vector of operation-state records bj, bj =
(oᵢ, sᵢ) is the iᵗʰ operation-state record, where oᵢ ∈ O is the user’s real operation and sᵢ ∈ S is the target state after performing oᵢ.

Table 2. User’s operations classes

<table>
<thead>
<tr>
<th>Behavior Event Class</th>
<th>Information Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>O_Click</td>
<td>(x, y): as integer data type</td>
<td>The operation records the coordinates (x, y) of mouse’s left-click action</td>
</tr>
<tr>
<td>O_Right-Click</td>
<td>(x, y): as integer data type</td>
<td>The operation records the coordinates (x, y) of mouse’s right-click action</td>
</tr>
<tr>
<td>O_Double-Click</td>
<td>(x, y): as integer data type</td>
<td>The operation records the coordinates (x, y) of mouse’s double-click action</td>
</tr>
<tr>
<td>O_Drag-and-Drop</td>
<td>(from_x, from_y), (to_x, to_y): as integer data type</td>
<td>The operation records the coordinates of drag (from_x, from_y) and drop (to_x, to_y) of mouse’s drag-and-drop action</td>
</tr>
<tr>
<td>O_Hot-Key</td>
<td>key: as string data type</td>
<td>The operation defines the hot-key performed by users.</td>
</tr>
<tr>
<td>O_Key-In</td>
<td>text: as string data type</td>
<td>The operation records the text keyed in the text area.</td>
</tr>
<tr>
<td>O_Choose-Menu-Item</td>
<td>text: as string data type</td>
<td>The operation records the text selected from the drop-down list.</td>
</tr>
</tbody>
</table>

Example 1: SO-FSM of the task “Set the font bold”

Figure 1 shows a SO-FSM example for the task: “Set the font bold” in MS Word® application software, where the starting state s₀ is attached with a screenshot image and the objective is to ask users to set the text in MS Word to “bold.” Thus, in the first step, either the drag-and-drop (e₁) or “hotkey ctrl+A” (e₂) method is used to select the texts to lead to the next state s₁, where the s₀ image is replaced by the s₁ image, but other unapplicable operations will be regarded as the failed operations (e₁) and then to the s₁₀ failed-state page, which indicates the wrong operations. After selecting text at s₁, users can use the sequence of menu items, i.e., “Format” (e₃) at s₁, “Font” (e₄) at s₂, “Font Style” (e₅) at s₅, and then “OK” (e₆) at s₆, to set the font to bold or to use the menu item, “Bold Type” (B) (e₇) at s₇, to achieve the same effect. Other inapplicable operations are regarded as failed operations (e₈, e₁₀, e₁₁, e₁₂). A special failed operation (e₁₃) to change the font size in the font menu will go to another failed-state page to show the misunderstanding status of users. Finally, users can accomplish the task objective if they can reach the final state, s₉.

A student’s SOP is illustrated as bold lines in Figure 1, where s/he first uses the hot-key, “Ctrl+a,” to select all of the text in a text area over the screenshot image at starting state s₀, and then the s₀ state moves to s₁ due to the transition input effect e₁. Subsequently, the student clicks the sequence of menu items, i.e., “Format” (55,12), “Font” (55,17), and “Size” (105,32), which match the input events e₃, e₄, and e₁₃, respectively. Finally, the failed-state s₉ is reached. The following SOP records the student’s four operations:

\[ SOP = <(\text{Hot-Key} \text{(Ctrl+a)}), s₁), (\text{Click}(55,12), s₂), (\text{Click}(55,17), s₃), (\text{Click}(105,32), s₉)> \]

According to the SOP, the SOS of a student can be automatically evaluated and assessed by the online system, and teachers are able to easily diagnose why and where a student misunderstood the skill of setting the given font to bold.

The construction of the simulation-based test based on SO-FSM

The SO-FSM scheme can be constructed by means of designing the desired SOP with required states (S), transitions (Σ), and associated input events (Δ defined in Table 1). Therefore, a graphical authoring tool has been designed and developed to facilitate the construction of the simulation-based test for assessing the SOS. Figure 2 describes the three construction steps used to create simulation-based tests based on the SO-FSM scheme (hereafter referred to as “SO-FSM-based tests”). By means of the SO-FSM-based graphical authoring tool, teachers can first visually design the SOP structure of an SO-FSM-based test with its question (or task) description and metadata (step 1 in Figure 3), and then edit the content of each state (sᵢ), consisting of a corresponding screenshot image, description, and metadata (step 2 in Figure 4), with the setting of associated input events (step 3 in Figure 4). Figure 5 shows the details of the SOP design structure and its states with associated input events of a SO-FSM-based test.
Figure 1. SOP of the “Set the font bold” task based on the SO-FSM

Figure 2. Three construction steps of creating a simulation-based test based on SO-FSM
Figure 3. Screenshot of designing the SOP structure of an SO-FSM-based test using a graphical authoring tool

Figure 4. Screenshot of editing a state (s_i) with associated input events using a graphical authoring tool
Example 2: Assessing the operations based on the SOP

The SOP recording the operation path is, thus, able to be used to automatically evaluate and assess the user’s misconceptions and learning problems via the online system. Accordingly, based on the task “set the font bold” in Figure 1, the testing results and assessment reports of three users with different MS Word abilities are illustrated to introduce the application of the online assessment mechanism. In Figure 6, (1) the “Novice User” could not perform the correct operation to select the target text, so a failed-state was achieved. (2) The “Confused User,” who confused the concept skill related to “can set font bold” with “can set font size,” could use the hot-key to select the text, but failed to make the selected text bold because s/he clicked the menu to change the font size instead. Thus, s/he went to
the failed-state $s_7$ along the transition $e_{12}$. (3) The “Passed User” was successful in selecting the text and making the text bold by using the bold menu, so the successful state was achieved.

The SOPs of three users are shown in Table 3, where both the “Novice User” and the “Confused User” failed to complete this software operation task. Moreover, teachers could find from their SOPs that the “Novice User” could not achieve the goal because s/he lacked the prerequisite skill of selecting the targeted text, and the “Confused User” misunderstood the operations of making the font bold and changing the font size. Therefore, this evaluation, based on SOPs, can assist teachers in offering personalized supplementary instruction for various students.

**Table 3. Various SOPs of three users**

<table>
<thead>
<tr>
<th>Learner</th>
<th>Novice User</th>
<th>Confused User</th>
<th>Passed User</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOP</td>
<td>$(\text{Click}(55,12), s_6)$</td>
<td>$(\text{Click}(55,12), s_2)$, $(\text{Click}(55,17), s_3)$, $(\text{Click}(105,32), s_7)$</td>
<td>$(\text{Hot-Key}(\text{Ctrl}+a), s_f)$, $(\text{Click}(55,12), s_2)$, $(\text{Click}(55,17), s_3)$, $(\text{Click}(105,52), s_4)$, $(\text{Click}(155,103), s_5)$</td>
</tr>
</tbody>
</table>

**The extension of QTI (version 2.1) for operating tests**

The IMS QTI Specification is a popular and standard format to represent the questions and assessments based on extensible markup language (XML) for supporting the interoperability among assessment systems. Version 2.1 is the current version (QTI, 2012). Besides the traditional multiple-choice questions, some kinds of advanced interaction in questions were defined using the QTI Specification, such as clicking the hot-spot in a graph and manipulating drag-and-drop images. However, the current specification lacks the definitions in relation to the users’ operating events for supporting the operating test. Therefore, this paper further extended the definitions of the QTI Specification to be capable of representing the proposal of the simulation-based test format for assessing SOSs. As seen in Figure 7, a state in an SO-FSM can be described in an assessment item, and these assessment items can be aggregated and controlled in an assessment test. Therefore, an assessment test can represent an SO-FSM. Among the assessment items, an `operationInteraction` element, including a prompt and an image to show the description and screenshot image respectively, is extended to capture users’ operating events, listed in Table 2. In order to map the captured events to the input events of an SO-FSM, an `eventMapping` element was extended from the `areaMapping` element in the QTI Specification, where each `eventMapEntry` element describes an input event belonging to the state. For
example, the state $s_0$ in Figure 1 must capture three input events: Drag-and-Drop in $e_1$, Hot-Key in $e_2$, and an undefined event in $e_3$. The first two events are defined in the eventMapEntry element with the information described in Table 1, and the undefined events are defined in defaultValue. Finally, the mapping identifier is assigned to the outcomeDeclaration element for finding the next state.

The transition rules in the SO-FSM can be described in an assessment test using the precondition constraints of the referred assessment items. As shown in Figure 8, the state $s_1$ in Figure 1 can be achieved if the outcome of $s_0$, having an identifier “qItem0,” is either $e1$ or $e2$.

**Figure 7.** Extending the definitions of events and operations in an assessment item

**Figure 8.** Describing the transition rules in an assessment test

### Implementation and experiment

#### Prototypical simulation-based software operation test system of an SO-FSM scheme

In order to evaluate the efficacy and effectiveness of the proposed SO-FSM scheme for assessing SOSs, a Simulation-based Software Operation Test System, called ShsOT system, has been developed to assist teachers in constructing the SO-FSM-based tests using the developed graphical authoring tool for students to practice and examine. Then, the system, via online resources, automatically assesses the students’ testing records based on SOPs. Figure 9 shows the system architecture of the ShsOT system, where teachers can connect to a designer server to
construct the desired SO-FSM-based questions and assessments. A student can register an account, login, take an issued assessment, and then receive an assessment report immediately by the evaluation server of the SbSOT system. Finally, the students’ SOPs can be recorded in an SOP database for further analysis by teachers.

The test interface of the SbSOT system can perform the SO-FSM-based tests and be run by a Web browser, e.g., Internet Explorer (IE), for students. As shown in Figure 10, students can read the question/task description: “Edit the text [Market News] to (1) Standard Kai Font, 36pt, Letter Spacing 3pt, and Shadow Effect” and (2) “Set to Neno-light Animation Effect.” Then students can virtually perform their operation in the operating area of the test interface. Figure 11 illustrates and exemplifies the online assessment procedure, wherein a student logs in to the SbSOT system (Figure 11.a), takes an SO-FSM-based test (Figure 11.b), virtually performs software operations using a Web-based test interface (Figure 11.c to Figure 11.g), and then instantly gets an assessment report generated by the automatic assessment process of the SbSOT system (Figure 11.h).

![System architecture of the SbSOT system](image)

**Figure 9.** System architecture of the SbSOT system

![The user test interface of the SbSOT system run on the IE Web browser](image)

**Figure 10.** The user test interface of the SbSOT system run on the IE Web browser
Experiment plan and execution

The design of an MS Word learning course

To evaluate the teaching and learning efficacy of the SbSOT system, five elementary school teachers, who have the experience of teaching MS Word software, were invited to participate in the design of the MS Word learning course. The course design included learning topics of SOSs, teaching materials, and SO-FSM-based content for practices and
tests created by the SbSOT system. The designed MS Word learning course was going to be used for 38 sixth grade elementary school students in Taiwan. Therefore, six topics of this learning course were selected by the teachers to meet the abilities of the targeted students, i.e., C1: Set Paragraph Attributes, C2: Insert and design Table, C3: Search and Replace, C4: Add Clip Art, C5: Set Font and Format, C6: Insert Symbols.

The design of the experimental learning activity

Thirty-eight sixth grade elementary school students in two classes, having similar SOSs, were defined as control and experimental groups taught by the same teacher chosen from five of the aforementioned participating teachers. As seen in Figure 12, the learning activity in this experiment, consisting of an introduction and four learning topics, was performed during classes consisting of 40 minutes/per week for 4 weeks. In the introductory part of the first week, the SbSOT system was introduced to all students in two groups. They were allowed to practice the SO-FSM-based test in relation to C5 and C6 topics because the pre- and post-tests of each week’s course were going to be conducted using the SbSOT system for convenience and experimental purposes. Afterwards, each week, the learning course composed of four topics was conducted, which consisted of a pre-test, lecture learning with practice exercises, and a post-test. The pre- and post-tests were conducted to evaluate the SOS abilities of all students before and after learning, respectively. In the learning portion of the course, the topics C1, C2, C3, and C4 were taught to students in both the control and experimental groups by means of a lecture followed by practice. The main difference between two groups is that the experimental group was taught not only via lecture with MS Word software, but also through learning with SbSOT practice. For instance, the teacher described the learning materials using real MS Word software, and then used the SO-FSM-based content to easily demonstrate the steps and explain the operational results through the automatic assessment of the SbSOT system. More importantly, students in the experimental group can conveniently engage in self-practice and self-reflection of the learning topics with the assistance of the SbSOT automatic assessment. The students in the control group self-practiced using the real MS Word software only and the assistance of the teacher.

![Figure 12. Experimental design and procedure](image)

Experimental results and analysis

Learning outcome analysis

The score distribution in Figure 13 and Table 4 shows the pre- and post-tests statistical results of two groups concerning the learning outcome of four topics, where the average progressive score (the difference between pre- and post-tests) of the experimental group (EG) is higher than control group (CG), that is, about 18.0925.

A paired-sample t-test is used to compare the pre-test and post-test scores, as shown in Table 5. Both in the experimental group and control group, the scores of four topics between the pre-test and post-test reach a significant difference, and the post-test scores (average in EG = 71.685 and average in CG = 60.26) are higher than those of the pre-test (average in EG = 39.2925 and average in CG = 45.96), respectively. This result reveals that the learning performances of the two groups have significantly improved.
Table 4. Summary Statistics of pre- and post-tests of four learning topics

<table>
<thead>
<tr>
<th>Group</th>
<th>Topic</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group (EG)</td>
<td>Pre-Test</td>
<td>41.41</td>
<td>48.42</td>
<td>40.60</td>
<td>26.74</td>
<td>39.2925</td>
</tr>
<tr>
<td></td>
<td>Post-Test</td>
<td>76.33</td>
<td>81.05</td>
<td>80.15</td>
<td>49.21</td>
<td>71.685</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>34.92</td>
<td>32.63</td>
<td>39.55</td>
<td>22.47</td>
<td>32.3925</td>
</tr>
<tr>
<td>Control Group (CG) (N = 19)</td>
<td>Pre-Test</td>
<td>47.37</td>
<td>54.60</td>
<td>50.22</td>
<td>31.65</td>
<td>45.96</td>
</tr>
<tr>
<td></td>
<td>Post-Test</td>
<td>61.43</td>
<td>66.44</td>
<td>72.81</td>
<td>40.36</td>
<td>60.26</td>
</tr>
<tr>
<td></td>
<td>Difference</td>
<td>14.06</td>
<td>11.84</td>
<td>22.59</td>
<td>8.71</td>
<td>14.3</td>
</tr>
</tbody>
</table>

Figure 13. Score distribution in pre- and post-tests

Table 5. Paired-sample t-test of pre- and post-test scores

<table>
<thead>
<tr>
<th>Group</th>
<th>Topic</th>
<th>Mean</th>
<th>Std. Deviation (SD)</th>
<th>t</th>
<th>p (Sig.) (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group (EG) (N = 19)</td>
<td>C1</td>
<td>-34.923</td>
<td>33.096</td>
<td>-4.599</td>
<td>0.000***</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>-32.6316</td>
<td>27.3414</td>
<td>-5.202</td>
<td>0.000***</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>-39.54789</td>
<td>25.87163</td>
<td>-6.663</td>
<td>0.000***</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>-22.472</td>
<td>23.501</td>
<td>-4.168</td>
<td>0.001***</td>
</tr>
<tr>
<td>Control Group (CG) (N = 19)</td>
<td>C1</td>
<td>-14.055</td>
<td>24.692</td>
<td>-2.481</td>
<td>0.023*</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>-11.8421</td>
<td>18.7979</td>
<td>-2.746</td>
<td>0.013**</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>-22.59421</td>
<td>26.19929</td>
<td>-3.759</td>
<td>0.001***</td>
</tr>
<tr>
<td></td>
<td>C4</td>
<td>-8.715</td>
<td>14.369</td>
<td>-2.644</td>
<td>0.017*</td>
</tr>
</tbody>
</table>

Table 6. ANCOVA for post-test scores by group

<table>
<thead>
<tr>
<th>Topic</th>
<th>Source</th>
<th>Sum of Squares (SS)</th>
<th>df</th>
<th>Mean Square (MS)</th>
<th>F</th>
<th>p (Sig.) (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Pre-test</td>
<td>14327.607</td>
<td>1</td>
<td>14327.607</td>
<td>22.265</td>
<td>0.000***</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>3158.109</td>
<td>1</td>
<td>3158.109</td>
<td>4.908</td>
<td>0.033*</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>22522.189</td>
<td>35</td>
<td>643.491</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>Pre-test</td>
<td>9920.478</td>
<td>1</td>
<td>9920.478</td>
<td>29.614</td>
<td>0.000***</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>2998.212</td>
<td>1</td>
<td>2998.212</td>
<td>8.950</td>
<td>0.005**</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>11724.917</td>
<td>35</td>
<td>334.998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>Pre-test</td>
<td>5623.898</td>
<td>1</td>
<td>5623.898</td>
<td>19.856</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>1128.664</td>
<td>1</td>
<td>1128.664</td>
<td>3.985</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>9913.214</td>
<td>35</td>
<td>283.235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>Pre-test</td>
<td>32526.091</td>
<td>1</td>
<td>32526.091</td>
<td>85.221</td>
<td>0.000***</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>1678.040</td>
<td>1</td>
<td>1678.040</td>
<td>4.395</td>
<td>0.043*</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>13362.547</td>
<td>35</td>
<td>381.787</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05. ** p < 0.01. *** p < 0.001.

In order to compare the improvement between the experimental and control groups, a one-way Analysis of Covariance (ANCOVA) is conducted, where the independent variable is the participants’ group, the dependent variable is participants’ post-test scores, and the covariate is participants’ pre-test scores. As shown in Table 6, with
the exception of the C3 topics, there are significant differences between two groups in terms of C1 \((F = 4.908, p < 0.05)\), C2 \((F = 8.950, p < 0.01)\), and C4 \((F = 4.395, p < 0.05)\) topics. The result reveals that the learning performance of the experimental group outperforms that of the control group.

**Questionnaire analysis**

After conducting the experimental learning activity, the questionnaires for teachers (as shown in Table 7) and students (as shown in Table 8) had also been defined to collect the degrees of satisfaction in terms of teaching and learning.

- **Analysis of teachers’ feedback**

According to the statistical results of the feedback received from five participating teachers (see Table 7 and Figure 14), the total average reaches a higher score of 4.17 (a score of 3.0 is neutral). This result indicates that teachers were highly satisfied with the functions of the SbSOT system. The teachers’ feedback indicates that the online assessment performed by the SbSOT can automatically evaluate operational records (a score of 4.6 in QT4), which not only saves teachers an evaluation cost (a score of 3.8 in QT1), but it can also assist teachers to understand students’ learning problems (a score of 4.0 in QT5) and further improve their teaching strategies (a score of 4.2 in QT6). This was according to the personalized assessment report. Consequently, teachers were very satisfied (a score of 3.8 in QT7) and highly agreed that the SbSOT system is useful and efficient for executing software operating tests (a score of 4.2 in QT2). The teachers were willing to continuously use the SbSOT system in MS Word teaching and assessment (a score of 4.6 in QT3).

**Table 7. Questionnaire of teachers’ satisfaction with using the SbSOT system (five-level Likert scale from 1 (strongly disagree) to 5 (strongly agree))**

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>QT1</td>
<td>Using the SbSOT system to online automatically assess students’ SOS can save the evaluation time.</td>
<td>3.8</td>
</tr>
<tr>
<td>QT2</td>
<td>The SbSOT system is suitable and efficient to perform software operating tests.</td>
<td>4.2</td>
</tr>
<tr>
<td>QT3</td>
<td>I am willing to continue using the SbSOT system to conduct software operating tests for students.</td>
<td>4.6</td>
</tr>
<tr>
<td>QT4</td>
<td>It is good to use the SbSOT system to perform online teaching and automatic assessments.</td>
<td>4.6</td>
</tr>
<tr>
<td>QT5</td>
<td>Students’ SOPs recorded in the SbSOT system can help me understand their learning status.</td>
<td>4.0</td>
</tr>
<tr>
<td>QT6</td>
<td>Students’ SOPs recorded in the SbSOT system can help me improve the teaching strategy.</td>
<td>4.2</td>
</tr>
<tr>
<td>QT7</td>
<td>Overall, I am satisfied with the SbSOT system.</td>
<td>3.8</td>
</tr>
<tr>
<td><strong>Total Average:</strong></td>
<td></td>
<td><strong>4.17</strong></td>
</tr>
</tbody>
</table>
• **Analysis of students’ feedback**

As stated in Table 8, the questionnaire for students consists of (1) a learning assistance (QS1 to QS5) aspect: assists in learning MS Word software in class, and (2) an assessment assistance (QS7 to QS10) aspect: assists in assessing students’ SOS abilities online in order to evaluate students’ satisfaction and feedback concerning the use of SbSOT system. As previously mentioned, the SbSOT system was used as the online assessment tool to evaluate the students’ pre- and post-test learning outcomes, so all students in both groups completed the questionnaire. Therefore, according to the statistical results shown in Table 8 and Figure 15, the students’ overall degree of satisfaction gains a satisfaction score of 3.713 (a score of 3.0 is neutral). To investigate the details, regarding the learning assistance aspect, the total average is a moderate satisfaction score of 3.554, where students were highly satisfied with the simulation-based MS Word learning support (a score of 4.2 in QS2), and were highly willing to recommend the SbSOT system to others (a score of 4.4 in QS5). However, the degree of satisfaction concerning operational similarity compared with the real MS Word (a score of 3 in QS1), the learning problems analysis (a score of 3.11 in QS3), and interest in using the system compared to using a traditional lecturing approach (a score of 3.06 in QS4) only received a slightly acceptable score. This evidence shows that although students consider the SbSOT system to be useful for learning MS Word, it still needs to be improved. Regarding the assessment assistance aspect, the total average is a satisfaction score of 3.872, greater than the score received for the learning assistance aspect, where students highly agreed that online simulated-based tests using the SbSOT system is much better than using pencil-and-paper tests (a score of 4.6 in QS6) because the evaluation results can be instantly received after assessment (a score of 4.3 in QS9). Although the ability of the SbSOT system to assess learning problems garnered only an acceptable score (a score of 3.0 in QS7 and a score of 3.06 in QS8), overall, students were highly satisfied with the assessment assistance of the SbSOT system (a score of 4.4 in QS10). The experiments and control groups provided similar feedback because the mean differences of satisfaction between them are very minute.

**Table 8. Questionnaire of students’ satisfaction for the SbSOT System (five-level Likert scale from 1 (strongly disagree) to 5 (strongly agree))**

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Item</th>
<th>Description</th>
<th>EG</th>
<th>CG</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Assistance (LA)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QS1</td>
<td></td>
<td>To operate the simulated MS Word created by the SbSOT system is the same as operating the real MS Word.</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>QS2</td>
<td></td>
<td>It is easy for me to learn operating steps of MS Word by means of the SbSOT system.</td>
<td>4.4</td>
<td>4</td>
<td>4.2</td>
</tr>
<tr>
<td>QS3</td>
<td></td>
<td>The SbSOT system can help me identify my learning problems in the MS Word course.</td>
<td>3.22</td>
<td>3</td>
<td>3.11</td>
</tr>
<tr>
<td>QS4</td>
<td></td>
<td>Learning with the SbSOT system is more interesting than learning with a traditional lecture approach only.</td>
<td>3.12</td>
<td>3</td>
<td>3.06</td>
</tr>
<tr>
<td>QS5</td>
<td></td>
<td>I will recommend other students to learn MS Word by using the SbSOT system.</td>
<td>4.6</td>
<td>4.2</td>
<td>4.4</td>
</tr>
<tr>
<td><strong>Total Average of Learning Assistance:</strong></td>
<td></td>
<td></td>
<td>3.668</td>
<td>3.44</td>
<td>3.554</td>
</tr>
<tr>
<td><strong>Assessment Assistance (AA)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QS6</td>
<td></td>
<td>MS Word assessment using the online simulation-based test by the SbSOT system is better than using a pencil-and-paper test.</td>
<td>4.7</td>
<td>4.5</td>
<td>4.6</td>
</tr>
<tr>
<td>QS7</td>
<td></td>
<td>The SbSOT system can accurately determine my operational problems according to my online assessment result.</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>QS8</td>
<td></td>
<td>I can understand my learning problems through the online simulation-based test of the SbSOT system.</td>
<td>3.12</td>
<td>3</td>
<td>3.06</td>
</tr>
<tr>
<td>QS9</td>
<td></td>
<td>It is good to get evaluation results instantly after I finished the online MS Word assessment.</td>
<td>4.4</td>
<td>4.2</td>
<td>4.3</td>
</tr>
<tr>
<td>QS10</td>
<td></td>
<td>Overall, I am satisfied with online assessment of the SbSOT system.</td>
<td>4.3</td>
<td>4.5</td>
<td>4.4</td>
</tr>
<tr>
<td><strong>Total Average of Assessment Assistance:</strong></td>
<td></td>
<td></td>
<td>3.904</td>
<td>3.84</td>
<td>3.872</td>
</tr>
<tr>
<td><strong>Total Average of Overall Satisfaction:</strong></td>
<td></td>
<td></td>
<td>3.786</td>
<td>3.64</td>
<td>3.713</td>
</tr>
</tbody>
</table>
Discussion and limitation

Discussion of the SbSOT system

Participant teachers highly agreed that the online simulation-based tests created by the SbSOT system were useful for teaching and for providing learning assistance in the MS Word course (proven by a total average satisfaction score of 4.17 in Table 7), although the SbSOT system is still under development. Especially, the online assessment with an automatic evaluation function can efficiently help teachers conduct the examination of operational software and save evaluation time over using real MS word software. Moreover, according to the assessment reports of the SbSOT system, teachers have a convenient approach to efficiently understand students’ learning problems and to present further explanations, while it is difficult to do this using traditional lecturing approaches with real MS Word software. Automatically generated personalized assessment reports are able to sustain students’ interests and improve their motivation to understand learning problems. For example, some students tried the online simulation-based tests many times to correct their mistakes and to actively discuss concerns with teachers after reading their assessment reports. This aforementioned phenomenon can also be proven by students’ feedback, presented in Table 8, where to learn by simulation-based content (a score of 4.2 in QS2) and to get instant evaluation results after assessment (a score of 4.3 in QS9) gained much higher satisfaction scores in contrast to the slightly acceptable score of learning problem analysis and operational similarity compared with the real MS Word. Furthermore, students were more satisfied with the assessment assistance aspect over the learning assistance aspect with a score of about 0.318. The possible reason for this result is that the main goal of the SbSOT system is designed to perform online assessment, so the support concerning learning assistance needs to be further improved.

Limitation of the SbSOT system

Even though the SbSOT system is useful and beneficial for teachers and students both on the assessment and learning of operational software, there are still limitations of our proposed system based on an SO-FSM scheme in terms of the following four aspects.

Capability of the ease-of-use of authoring reconfigurable simulation-based tests

In the SbSOT system, a graphical authoring tool has been developed to assist users to efficiently construct a reconfigurable SO-FSM-based test. Participant teachers were able to learn to use a graphical authoring tool and create their-own SO-FSM-based tests after about four hours of training. However, teachers commented that the major problem and cost of the authoring tool is that the more complex the desired SO-FSM-based test is, the more difficult the authoring process is because they have to determine all possible operating paths and prepare the corresponding screenshot images of states in advance. Accordingly, the ease-of-use of an authoring tool must be
further improved. For example, the desired SOP with required stuff of SO-FSM-based tests can be automatically constructed by a real-time recording of the teacher’s operating process on the real software. Thereby, teachers can easily complete their SO-FSM-based tests after some revision. The fixed structure of SOPs also confines the flexibility of authoring SO-FSM-based tests; therefore, relieving this constraint via the flexible design, e.g., object-oriented approach, is another significant concern.

**Capability of operational software’s supportability**

Based on the reconfigurable simulation-based test scheme, an SO-FSM-based test can be easily simulated by corresponding software screenshots with customized functions, defined as event and operation classes in Table 1 and 2. Therefore, not only MS Word, but other operational software, like Web browsers, can also be simulated and created by the SbSOT system as long as these customized functions can meet the operational requirements. An example of “Change Internet Explorer Security Settings” is illustrated and exemplified in Figure 16. However, several operational processes cannot yet be supported and satisfied by current functional definitions, such as table formatting, image resizing, specific effects, etc. Accordingly, these event and operation class definitions have to be extended and developed to meet various requirements of operational software.

![Graphical SOP Structure](image)

**Figure 16.** An IE browser example for “Change Internet Explorer Security Settings” using the SO-FSM-based test

**Capability of learning problem diagnosis**

The personalized assessment report can be automatically generated by evaluating the recorded SOP. However, the current analysis process is able to identify the problematic operating steps and corresponding misconceptions only; the detailed learning problems concerning the cause and effect operational behaviors (with reasons and suggestions) cannot yet be assessed. This finding could be the reason explaining why students’ average satisfaction degree concerning learning problem analysis is only slightly acceptable (i.e., score of 3.11 in QS3, score of 3.0 in QS7, and score of 3.06 in QS8). Therefore, the capability of the diagnosis process needs to be further enhanced.
Capability of learning assistance functionality

The main goal of the SbSOT system is designed for online automatic assessment, so students cannot get any help and guidance until they complete the assessments. Consequently, it might be easy to impact students’ confidence and decrease their interest, although they use the SbSOT system to learn and practice. This could explain why the average score of learning assistance for students is 3.554, which is less than the 3.872 average score of assessment assistance. Accordingly, in order to improve the learning assistance aspect, it is necessary to provide students with learning guidance functionality during the online operation.

Conclusion

In order to address the problems with respect to constructing and conducting the simulation-based examinations for automatic online assessment of software operating skills, a reconfigurable simulation-based test scheme, called SO-FSM, was proposed. Furthermore, its system, called the SbSOT system, has also been implemented. Thereby, teachers can efficiently construct the simulation-based tests by editing the software operating process using the developed graphical authoring tool based on the proposed SO-FSM scheme. Students can virtually operate the SO-FSM-based tests via the Web browser and their operational behavior, defined as software operating paths, can be recorded and evaluated for the automatic generation of assessment reports. Students can then practice and reflect on the results of operational software learning via the SbSOT system’s supports. In order to support interoperability, the definition of QTI version 2.1 has been extended to meet the construction of SO-FSM-based tests as well. Furthermore, the pilot experimental results show that the learning performance of students in the experimental group using assistance from the SbSOT system significantly outperforms the results of the control group using a traditional lecture approach according to the ANCOVA analysis in terms of four pre- and post-tests. According to feedback, teachers highly agreed that the SbSOT system was useful for teaching and for providing learning assistance in an MS Word course, especially for the online assessment with automatic evaluations being used to identify students’ learning problems and to improve teaching strategies. Similarly, students were also satisfied with the assistance of the SbSOT system and prefer the online assessment to the online learning. Although the SbSOT system is useful and beneficial for teachers and students both on the assessment and learning of operational software, the existing limitations must still be addressed and solved. Therefore, in the near future, further concerns will be to enhance the ease-of-use, supportability, diagnostics, and functionality of the SbSOT system. Also, we will seek to break the constraint of predefined whole possible operating paths of a test, to support more various processes of operational software (e.g., table formatting) to diagnose learning problems in depth, and to offer learning guidance during online operation.

Acknowledgements

This research was partially supported by National Science Council of Republic of China under the numbers of NSC 101-2511-S-024-004-MY3 and NSC 102-2511-S-024-009-MY2.

References


A Study of Perceptual Typologies on Computer Based Assessment (CBA): Instructor and Student Perspectives

Jin-Young Kim
Ulsan National Institute of Science and Technology, UNIST-gil 50, Ulsan 689-798, Republic of Korea // jyk@unist.ac.kr

ABSTRACT
This study explores and describes different viewpoints on Computer Based Assessment (CBA) by using Q methodology to identify perspectives of students and instructors and classify these into perceptual typologies. Thirty undergraduate students taking CBA courses and fifteen instructors adopting CBA into their curriculum at a university in Korea, were chosen as participants in this study. As a result of the study, firstly, four types of learners were identified and given the following descriptive labels: (I) CBA Dissatisfaction Type, (II) CBA Friendly Type, (III) Adjustment Seeker Type, and (IV) CBA Preparative Type. Secondly, three types of instructors were classified and given the following descriptive labels: (A) CBA Preferred Type, (B) Supplementary Need Type, and (C) Yes-But Mixed Type. Educational and pedagogical implications of these results are discussed in the context of CBA course design.

Keywords
Computer-based assessment (CBA), Blended learning, Users’ perspectives, Q method, Typology

Introduction
Student assessment is considered an integral part of the learning process, which can be used for learning evaluation, student motivation and grading, and program adaptation (Emmanouilidou et al., 2012). The application of supporting technologies such as computer based assessment (hereafter referred to as “CBA”) has mainly been used in a blended learning environment in order to assess student learning, as well as to enhance students’ self-directed learning (McBain, 2001). CBA has become a crucial aspect of overall assessment strategy as well as blended learning for students (Waddell et al., 2012). The rationale for this strategy is that CBA increases and improves student’s participation and attendance (Marriott & Lau, 2008). Therefore, the application of CBA to various learning environments, and the issues involved with doing so, are the subject of much discussion.

CBA offers enormous prospects for innovation in learning, testing, and assessment such as self-evaluation, useful feedback, space and time flexibility, and multimedia usage. Paper based formative assessments are currently complementing, or have already been replaced by innovative and powerful CBA instruments these days (McDowell, 2002). Despite the numerous benefits just listed, paper-based evaluation is still commonly used in contemporary learning environments. (Hole-Elders et al., 2008). The enhancement of CBA will positively influence teaching effectiveness; therefore, diversified efforts should be made to establish CBA in the learning environment.

Despite many studies on the application or effectiveness of CBA, research on users’ perspectives toward CBA is still inadequate. CBA has been evaluated from the instructor’s perspective in the existing literature (e.g., Magliano and Graesser, 2012; McKenna, 2001), but research into learners’ viewpoints about CBA is limited. The effective development of CBA depends on both instructors’ and students’ acceptance. Consequently, there is a need for a more formal approach to understanding the perceptual perspectives of instructors and learners with CBA. The purpose of this paper is to first analyze the learners’ perceptual typology regarding CBA. In other words, this study purposes to investigate and categorize the learner’s perception of CBA, ultimately investigating the learner’s particular typology with regard to CBA. Secondly, the perception of professors and instructors who are currently conducting CBA was analyzed to investigate their perceptual typologies and compare the differences between instructors and learners. With the aim of providing a typological analysis of users’ perspectives, this study used Q methodology to identify different viewpoints on CBA.

Originally evolving from factor-analytic theory, Q methodology combines aspects of both qualitative and quantitative research traditions as a way to study the subjectivity involved in any situation (e.g., Kim, 2012). This method is ideal for deeply exploring areas of complex perceptions or opinions. Participants are asked to sort and rank a sample of statements concerning the subject of research (i.e., Q sorting). Then, the Q sorts are correlated and factor

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.
analyzed, resulting in different types that are qualitatively interpreted, providing accounts of understandings of the subject (Brown, 1980).

The result of this study partially contributes to the effective application of CBA when designing classes, as well as offering benefits on research objectives and settings. This study was conducted on full-time students taking blended learning courses and instructors actually performing CBA at a regular university, excluding part-time students who are taking 100% online courses.

Literature review

What CBA is

CBA utilizes computer technology, including the Web for assessing student learning (Bull & McKenna, 2000) and provides flexibility in style and content with reduced tutor dependency (Boud, 2000). It is also defined as a technique which “encompasses the use of computers to deliver, mark or analyze assignments or exams” (Sim et al., 2004, p. 217).

CBA can be categorized into formative and summative assessments. Formative assessment is designed to aid learning through familiarizing students with course contents, informing them about gaps in their learning, and providing feedback to guide the direction of learning. This occurs during the course of instruction. Providing formative assessment has been considered a significant benefit to student learning and leads to enhanced learning outcomes (Black & Wiliam, 1998; Rolfe & McPherson, 1995; Sadler, 2010). On the other hand, the goal of summative assessment is to evaluate student learning. Summative assessment measures what students have learned at the end of an instructional unit and is concerned with accountability and certification. It helps to establish whether students have attained the goals set for them. Summative assessments, which are generally high stakes, provide little opportunity for effective learning through feedback. (Economides, 2009; Ellery, 2008). So, CBA includes adaptive testing, analysis of the content of discussion boards, automated essay marking, delivery of exam papers, multiple choice testing, and so on (Sim et al., 2004).

CBA plays a strong role in overcoming the spatial limitations of formative assessment using multimedia, drawing automatic results, improving educational applications through the analysis of questions and evaluation results, providing immediate feedback, reducing errors in evaluation, increasing the speed of results, tracking students performance, and more (McKenna, 2000). However, its application in the learning environment still lags behind. This is because of a lack of technical knowledge and skills to develop CBA, a lack of the instructor’s time or desire to apply CBA, a lack of resources, and a lack of computer facilities to execute CBA (Hole-Elders et al., 2008).

Users’ perceptions of CBA

The previous studies on CBA have addressed many aspects of CBA: effective application strategies for CBA (Bull & McKenna, 2000; Hole-Elders et al., 2008), educational effects (Zakrzewski & Bull, 1998), quality assurance of CBA such as clarity, validity, and reliability (Conole & Warburton, 2005), viewpoints on CBA (McKenna, 2001; O’Hare, 2001; Waddel et al., 2012), interaction/feedback issues (Wang, 2007; Yorke 2003), and technical issues (Boyle & O’Hare, 2003; Magliano & Graesser, 2012). For example, ineffective tutor feedback in the early stage is being replaced by qualitatively superior electronic feedback (Yorke 2003). Regarding the effects of CBA, there is some evidence that formative CBA can improve final test results (Zakrezewski & Bull, 1999) but there is also a finding that frequent high stakes assessment has a negative motivational impact on learning (Harlen, 2005). In addition, proper computer hardware, and infrastructure are needed to effectively adopt CBA. Moreover, the establishment of a clear evaluation purpose, the security to prevent the external leakage of tests, the provision of proper feedback and training of instructors, the evaluation of CBA and the use of item banks have been suggested for the effective application of CBA (Hols-Elders, Bloemendaal, Bos, Quaak, Sikstermans, & Jong, 2008).

Regarding the perspective of professors and their opinions, the question format of CBA was the main issue. Student-constructed responses, such as essays and short-answer questions, can be useful tools to gauge student learning outcomes and comprehension strategies. However, multiple-choice items remain the most widely and commonly
used item format in CBA due to the difficulty of grading for essay format and differences in criteria that humans versus computers use (Magliano & Graesser, 2012). The application of various evaluation methods, other than multiple-choice questions, that enable instructors to measure the higher level skills of students, such as synthesis or analysis, have not yet been adopted by CBA (Conole & Warburton, 2005). Of course the development of high-quality multiple choice items is still a difficult skill to acquire (Boyle & O’Hare, 2003); the objective style testing within CBA is well established for standardized assessment across all levels of education (Bennett, 2002). However, Nicol & Macfarlane-Dick (2006) strongly claimed the theoretical basis of CBA is still inadequate and the concept of assessment is still far behind.

Instructors regard computer based summative assessment as a means of improving students’ participation and attendance, and also as key to developing self-regulation and independence in the education experience (Marriott & Lau, 2008; Novakovich & Long, 2013). Similarly, McKenna’s study (2001) indicated that most instructors applying CBA to their classes are using it for the purpose of performing a formative assessment and they have said the reason they adopted CBA is to promote student’s regular learning. The instructors revealed that the improvement of self-driven learning improves students’ interaction with the instructor during class and allows students to deeply study the concepts they could not understand from CBA. Time savings for large class sizes and the flexibility of time and space for students were other reasons to adopt CBA. Several reasons were given for not adopting CBA, including its unsuitability for discursive courses, its inability to assess problem-solving skills, its question design limitations, and department regulations preventing CBA for final tests. The instructors voiced their concern over CBA’s mechanistic approach to learning that encourages students to care more about the correct answer than the knowledge they should learn. CBA was rather suitable for evaluating basic knowledge and comprehension. Also, the instructors found it difficult to create CBA questions and update new questions; they mentioned the need for staff development in question design or, if possible, utilization of question banks (McKenna, 2001).

The use of CBA in humanities is limited compared to other disciplines. In the Computer-assisted Assessment (CAA) Centre national survey, humanities users have focused on CBA’s role as supplementary material, not key material (McKenna, 2000). In a study conducted by Broughton et al. (2013), instructors applied CBA to their classes for the following key reasons: saving time, enhancing students’ responsibility, departmental influence, and inheritance from existing modules. The instructors recognized CBA as an effective tool in assessing large groups of students but they also pointed out their difficulty in determining whether students really completed the tests on their own or by engaging in plagiarism. In order to overcome the drawbacks of CBA and its limitations in measuring the in-depth understanding of students, it has been reported that the instructors have used other assessment techniques such as coursework or projects as supplementary methods and reduced the weight of computer based summative tests.

Further, CBA should look for ways to prevent cheating in online assessments. Methods such as blinker screen invigilation, interleaving participants using different tests, and randomizing item and response orders have been proposed (Bull & McKenna, 2000; Conole & Warburton, 2005). In addition to the cheating issue, the security issue is regarded as an essential issue in CBA. Now it is hard to protect security since the test data is being sent across the internet. Therefore, in order to ensure that individuals cannot access unauthorized materials, encryption techniques can be used to ensure the security of the questions and answers. Examinations can be loaded onto the server at the last minute (Sim et al., 2004).

Lastly, regarding the administrative support, Boyle & O’Hare (2003) claimed that most professors and instructors are responsible for creating CBA questions on their own; however, the training or education for them to produce more clear and reliable questions is significantly inadequate. In addition, adequate technical and administrative support from school staff is needed to apply CBA to classes.

Compared to the perceptions of professors, the studies on CBA from the learner’s perspective are limited. In terms of students’ viewpoints, just like the use of CBA is initially time consuming for faculty, students recognize online classes are more time consuming and require more effort compared to other traditional classes (Knight et al. 1998 cited by Liu, 2012). However, students considered CBA to offer relatively fairer assessment than paper-based tests. The immediacy of the feedback and the relaxed way of learning are the key advantages of CBA (McKenna, 2001). The inclusion of high quality graphics was cited as an advantage of CBA by students in a study about student views of CBA. The majority of respondents answered “neutral (neither agree or disagree)” to the question whether CBA is more challenging compared to traditional classes, while showing a fairly equal number of respondents who either agreed or disagreed on the same question (O’Hare, 2001). In another empirical study, about 70% of students...
responded that CBA is the one of the best methods of examining students and they were also highly and moderately acceptable to apply CBA in their learning. 90% of students felt that CBA did not disadvantage them even in a high-stakes situation (Escudier et al., 2011).

In a focus group session, students explained that having a series of low stake assessments was beneficial to them. That is, the tests are a study aid helping them to build up knowledge and read the course materials in small chunks without too much stress. The students stated that with CBA it was more relaxing to do the test at their own pace than it would be in an exam context. However, they suggested more detailed feedback on their answers, especially if they got them wrong (Waddel et al., 2012). This perception of the lack of effective feedback in CBA is in line with instructors’ perception in a study by Broughton et al. (2013). In the Broughton et al. study, instructors felt that more improved individualized feedback, along with general comments should be provided to students in CBA. As Black and William (1998) suggested, giving proper feedback to guide the direction of learning is essential in CBA, especially formative assessments. If a CBA database system is equipped with well-designed feedback data, it can provide learners with more effective feedback and facilitate learning (Wang, 2007). However, some studies have raised a question regarding the quality of feedback to students in CBA. For instance, compared to paper based tests, some academic staff members felt there was a lack of individual comments in CBA with a lengthy written answer format. The process of typing text in a textbox on screen caused a slowdown in grading thus was not used (Sheader et al., 2006). Similarly, instructors commented that one of the disadvantages of CBA was the feedback to students because it is hard to provide meaningful feedback in analysis or synthesis of information (McKenna, 2001). Another empirical study on attitudes of CBA feedback conducted by Wong et al. (2001) found that students perceived that feedback or explanations provided by web-based assessment system were useful, however, some respondents suggested that more detailed feedback should be given.

Terzis and Economides (2011) created the following categories to investigate students’ intentions while studying the students’ behavioral intention to use CBA: perceived usefulness (e.g., using CBA will improve my work), perceived ease of use (e.g., my interaction with the system is clear and understandable), facilitating conditions (e.g., when I need help to use CBA, someone is there to help me), content (e.g., CBA’s questions were useful for my course), perceived playfulness (e.g., using CBA gives me enjoyment), behavioral intention to use CBA (e.g., I plan to use CBA in the future) and so on.

Study context

This study was conducted at the Ulsan National Institute of Science and Technology (UNIST) in Korea. UNIST has adopted a blended approach combining online and offline instruction to meet the needs of educational efficiency since 2010. The university employs the term “e-Education,” to refer to the self-driven and intensive learning support system based on the Blackboard learning management system (LMS). UNIST is the only school to adopt the Blackboard system in Korea. In order to support e-Education, the Center for Teaching and Learning at UNIST provides training, workshops, technical maintenance, and support. The center leads and encourages blended course. There have been blended course offerings implementing CBA since 2010; these are mainly large enrollment courses for first-year students including basic science, math, IT, arts, humanities, and social science courses (e.g., Practical IT, Statistics, Communication, Philosophy, Music). By 2013, approximately 15% of the total course offerings are now blended learning courses. These courses are designed and taught by individual faculty members, and facilitated by assistant instructors under the supervision of a professor.

Blended learning classes reduce face-to-face time in the classroom by 30% to 50%. That is, the courses at UNIST incorporate online classes as part of the course where students engage in independent study each week. Accompanying classroom lessons, these online sessions require learners to perform the following tasks mostly via Blackboard: browse the Internet, view online lectures or streaming media content, read various course materials, participate in asynchronous online discussions, take online quizzes and tests, engage in group problem-solving and collaborative tasks, share content and perform peer evaluations, submit written assignments, and receive feedback on those assignments. The university provides free WiFi wireless technology on campus where most students reside, and enables students to access the Blackboard LMS by using the iPhone, thereby facilitating Internet access and constant online connectivity.
The professors at UNIST who are applying CBA to their classes utilize the low-stake summative assessment, scored formative online end of chapter quizzes, and unscored practice tests. The weight of CBA scores is not that significant; most of them range from 10% to 25% of the total grade. That is, CBA is used as a complementary tool rather than a substitute tool for student evaluation. For IT courses, the professors apply CBA after finishing each unit and use the Adaptive Learning Environments Model (e.g., Wang and Walberg, 1983) that allows those students reaching at least 80% of academic achievement in assessment to move on to the next unit. For those who are evaluated as inadequate to move on, individual supplementary instruction is provided.

**Methodology**

**Research question**

This study is to investigate the user’s perception and attitude of CBA and the typology of these perceptions. To do this, Q methodology was used. Q methodology differs from R methodology (survey and questionnaires) in that the latter asks respondents to express views on isolated statements. The Q methodology, instead, is a research method used to identify respondents' views on particular statements in the context of the valuation of all statements presented. Scholars who rely on Q methodology criticize the use of predetermined scales in ordinary R methodology studies. Furthermore, as opposed to R methodology, Q methodology traditionally intends to give a picture of the perspectives that exist among the population, rather than analyzing the level of support for those perspectives among the population. This implies that the procedure for sampling respondents is usually different from that in R methodology. Rather than random sampling with large sample sizes, Q methodology relies on purposive sampling with smaller sample sizes (Kim, 2008).

This study sought answers to the following research questions.

Q 1. What types can be discerned in students’ perceptions of CBA and what are the characteristics of each type?

Q 2. What types can be found in instructors’ perceptions of CBA and what are the characteristics of each type?

**Q statements (Q sample)**

The procedure of Q methodology comprises several stages. The first stage is defining the concourse, that is, the full range of discourses on the particular issue under study. For derivation of the concourse, preliminary research was conducted through a review of the academic literature on CBA, a focus group, informal conversation, various sources including oral and written opinions, and reflection on personal experiences. The focus group consisted of three students and two faculty members who were asked to discuss freely their experiences with CBA. A total of 94 statements for students’ views and 86 statements for instructors’ views were derived from the preliminary research. The statements were checked for duplication, comprehensibility, and similarity. Then, in order to reduce the statements to a manageable number, a Q sample which was used in the main study, was drawn from the concourse. The Q sample consists of 42 statements for students and 33 statements for instructors, selected for their representativeness and suitability in reflecting the diversity of statements in the concourse (refer to Table 1 and 2). These statements were self-referent, that is to say, they were expressions of opinion rather than factual descriptions, so that a respondent could project him- or herself through them.

**Participants (P sample)**

The next stage, which plays an important role in Q methodology, is the selection of a P sample, namely, respondents who perform the task of sorting Q statements and participate in in-depth interviews. Q methodology handles the difference of importance within individuals and not between individuals, so there is no limit to the size of the P sample; the size of the P sample can be different according to the purpose of the study or the number of factors, but it follows basic research design principles and small sample principles (Brown, 1980). Participants for students’ perspectives were recruited from the university. There was recruitment advertising for the research on campus and also an assistant researcher recruited participants from several large-scale classes. Among the volunteers, a total of 30 students was selected after considering gender and school years. The students were full-time students in various majors and from various years of enrollment at the university. The average age of the
student P sample was 21.3 (SD = 1.34); 63.3% were men, 36.7% women. Freshman participants constituted 36.7% (n = 11) of the sample, sophomores 30% (n = 9), juniors 23.3% (n = 6), and seniors 10% (n = 3).

Moreover, a purposive sample of 15 instructors who used CBA for their course was selected for this study. The respondents for instructors’ views included 8 full-time professors and 7 assistant instructors. The assistant instructors in this school are not graduate students; they are full-time faculty who are hired to assist the main instructors with teaching. Their career in education ranged from 1 to 34 years, with a mean length of 6.4 years. The subjects they are teaching are basic science courses (math, chemistry, biology, IT, statistics and computer programming) and humanities courses (communication, society and culture, literature, management, philosophy and arts).

Data collection (Q sorts)

At each site where participants were interviewed, a large table was used for the Q sorting task. Participants received instructions and familiarized themselves with the statements regarding CBA printed on index cards. Then, firstly, each respondent (in the P sample) performed the Q sorting task. This task involves sorting well-composed statements about CBA in the Q sample after reading them, by using a compulsive classification method. The classification method required respondent to sort the statements on a scale ranging from “most agreed” (+4) to “most disagreed” (-4); the participants in the survey modeled their cognition by themselves. Once finished, respondents participated in an in-depth interview about CBA to provide a fuller understanding of the characteristics of each type. That is, an investigator conducted a post-sort interview offering participants the chance to offer additional statements. In this process, the investigator questioned each participant about his/her Q-sort reasoning and further opinions. After the Q sorts were obtained through the sorting work, they were analyzed using the computer program QUANL. The in-depth interviews and Q sorts for this study were conducted during the fall semester in 2013.

Table 1. Q statements and z-scores of learners’ typology

<table>
<thead>
<tr>
<th>Q Statements</th>
<th>Z scores by Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I think CBA has a fairer grading system than the traditional test.</td>
<td>I</td>
</tr>
<tr>
<td>2. Currently CBA is mainly composed of multiple-choice questions but I think these multi-choice questions can’t exactly evaluate student’s ability.</td>
<td>.2</td>
</tr>
<tr>
<td>3. Diverse assessment methods, not only paper-based tests, but also problem-solving ability evaluation, team work evaluation, oral tests, and field application evaluation, should be needed and introduced in order to evaluate student ability.</td>
<td>1.3</td>
</tr>
<tr>
<td>4. CBA courses made me nervous throughout a semester due to their heavy assignments and lots of work to do.</td>
<td>.2</td>
</tr>
<tr>
<td>5. CBA needs to be improved in many aspects (e.g., diversification of question types).</td>
<td>.5</td>
</tr>
<tr>
<td>6. I don’t think CBA helps student-professor interaction during the class.</td>
<td>-.9</td>
</tr>
<tr>
<td>7. CBA is a good evaluation method to allow student flexibly to choose study time and place.</td>
<td>-.0</td>
</tr>
<tr>
<td>8. Anyone can take CBA regardless of computer utilization of each student. In other words, one’s skill at computers has no correlation with test results</td>
<td>-.2</td>
</tr>
<tr>
<td>9. CBA significantly helped me to understand and learn the courses.</td>
<td>-.1</td>
</tr>
<tr>
<td>10. There are subjects more suitable to CBA.</td>
<td>.8</td>
</tr>
<tr>
<td>11. CBA is very convenient because it allows students to check the incorrect answers and grades right after taking a test.</td>
<td>.0</td>
</tr>
<tr>
<td>12. CBA is a very innovative method. It should be more widely used in learning environment.</td>
<td>-.9</td>
</tr>
<tr>
<td>13. CBA gives a unique, interesting experience to me.</td>
<td>-.5</td>
</tr>
<tr>
<td>14. CBA is only good for mechanical learning, which leads me to select the right answer, rather than giving me an opportunity to learn what I really have to learn.</td>
<td>1.7</td>
</tr>
<tr>
<td>15. Assistant instructors or administrative support in CBA doesn’t really help me while</td>
<td>-.4</td>
</tr>
</tbody>
</table>
taking a test.

16. My anxiety became much less intense when taking a CBA that allows me to look up references and internet while taking a test compared to taking a test at the same time, in a real-time manner. .1 .0 1.1 1.6

17. Questions or instructions given in CBA are clear without any confusion. -.4 2.1 .7 .4

18. I am worried what if errors occur while taking a computer based test. For example, the answers I marked are not saved in the system or unexpected errors in the server that make my computer crash in the middle of a test. .7 -1.2 1.9 1.5

19. CBA should be used for practice tests rather than important tests or grading. .7 -.3 -.1 .3

20. I prefer more detailed feedback on answers, especially if I got them wrong. 1.0 0 1.1 2.0

21. I am mostly satisfied with the current technical environment of CBA tests including the network connection and computer utilization. -.2 1.9 -.5 -.2

22. I think CBA is a useful assessment method since I could take a test containing high quality graphics, images and videos. -.0 2.3 -.7 .4

23. CBA results may produce totally different results than paper-based tests, mid-terms or final examinations given once or twice a semester. .2 -.1 .4 -.8

24. I wish CBA became a supplementary method in evaluating students. 1.9 -.5 1.3 1.2

25. I guess CBA will be more widely used in the learning environment than the traditional test. 1.4 .3 -.1 -.3

26. I have trouble reading documents online. Therefore, I usually print the data and read. 1.4 .2 -.0 .2

27. I think I was more active when participating in CBA class than traditional classes. -.8 .2 -1.5 -1.1

28. I’m not satisfied with the IT-dependent tools, such as CBA or online classes. .2 -.6 -1.0 -.3

29. I would choose a paper-based test rather than CBA if I had an opportunity to choose from those two options. 1.7 -.5 -.6 .2

30. Face-to-face feedback from the instructor should be increased rather than the feedback given through the computer. 1.4 -.3 -1.3 .3

31. I think my responsibility or autonomy for learning has been improved through CBA. -2.2 1.4 -.7 -1.4

32. CBA should be included in more varied subjects compared to current status, including the core subjects. -1.2 -.2 .0 .0

33. I don’t think CBA is a useful and valuable activity in student evaluation. -.2 -1.6 -1.5 .1

34. CBA should be limited to basic courses, such as testing basic knowledge and terminology. 1.6 -.2 .5 -1.4

35. CBA didn’t require me to spend more time and effort compared to traditional classes. -2.0 -1.7 -1.8 .5

36. The weight of the CBA score out of the total points should be higher than now. -1.2 -.7 -1.9 -2.0

37. I don’t really care about CBA since the courses applying CBA are not major courses and their scoring weights are not that high. -.4 .2 .4 -1.0

38. Questions on CBA are too easy and simple. -.3 -.1 -.9 -.8

39. I don’t really care if it’s a paper-based test or CBA. Because I believe the results do not depend on the assessment method. -1.1 .8 -1.1 -.9

40. I am interested in all computer or internet based activities, including computer utilization training, class or test. -1.5 .3 -.1 -.4

41. I became less burdened by study because I could be assessed in each unit. -.4 .9 -1.0 .6

42. I think CBA allows students to engage in cheating much more easily. .9 .9 .5 1.6
Table 2. Q statements and z-scores of instructors’ typology

<table>
<thead>
<tr>
<th>Q Statements</th>
<th>Z scores by Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I would have a hard time operating CBA without administrative support from AI or faculty in the Center of Teaching and Learning.</td>
<td>A  -0.9</td>
</tr>
<tr>
<td>2. CBA saves my time on grading tests for large class sizes.</td>
<td>A  1.5</td>
</tr>
<tr>
<td>3. The most typical question type in CBA, the multiple-choice question, has no significant problem in evaluating students.</td>
<td>A -0.5</td>
</tr>
<tr>
<td>4. Not by recommendation from school or department, I initatively adopted CBA.</td>
<td>A -0.5</td>
</tr>
<tr>
<td>5. I don’t think a professor’s computer utilization affects the adoption of CBT.</td>
<td>A  0.7</td>
</tr>
<tr>
<td>6. I found no significant difference in test results between CBA and traditional tests.</td>
<td>A  0.4</td>
</tr>
<tr>
<td>7. I think my acceptance of blended classes or CBA is lower than other instructors.</td>
<td>A  0.2</td>
</tr>
<tr>
<td>8. CBA may become a mechanical activity or study for students focusing on providing the right answer rather than in-depth analysis or learning.</td>
<td>A -2.1</td>
</tr>
<tr>
<td>9. Currently CBA is a supplementary method for student evaluation but should be the main method in the near future.</td>
<td>A  0.6</td>
</tr>
<tr>
<td>10. CBA cannot really test synthesis to any great level so it should be improved</td>
<td>A  0.3</td>
</tr>
<tr>
<td>11. The questions type of CBA (multiple-choice question) is somewhat limited; therefore, new evaluation methods such as class activities or projects should be combined.</td>
<td>A  0.5</td>
</tr>
<tr>
<td>12. The feedback provided by CBA is much more useful than lecture notes or textbook.</td>
<td>A -0.3</td>
</tr>
<tr>
<td>13. It is hard to apply CBA to all courses. I think there are particular subjects suitable for CBA.</td>
<td>A -1.0</td>
</tr>
<tr>
<td>14. CBA is an effective assessment method that allows students taking test containing high quality graphics, images, and videos.</td>
<td>A  1.4</td>
</tr>
<tr>
<td>15. I don’t think CBA improved my interaction with students during the class.</td>
<td>A -1.4</td>
</tr>
<tr>
<td>16. Professors can perform CBA without timing or spatial limitation and this can help professors to managing classes.</td>
<td>A  1.4</td>
</tr>
<tr>
<td>17. CBA improves student’s participation and responsibility for learning.</td>
<td>A  1.3</td>
</tr>
<tr>
<td>18. CBA tailored to the characteristics of each student should be performed.</td>
<td>A -0.5</td>
</tr>
<tr>
<td>19. It is useful to operate class since CBA allows me to track student’s academic activities.</td>
<td>A -0.5</td>
</tr>
<tr>
<td>20. It takes a lot of time to make and update CBA questions; therefore, I don’t think CBA is an effective educational tool.</td>
<td>A -1.3</td>
</tr>
<tr>
<td>21. Resources to set CBA questions are not enough..</td>
<td>A  0.6</td>
</tr>
<tr>
<td>22. I would like to recommend adopting CBA for other colleagues.</td>
<td>A  1.1</td>
</tr>
<tr>
<td>23. CBA is a new paradigm in learning; personally, it is very interesting to me.</td>
<td>A  1.1</td>
</tr>
<tr>
<td>24. I think students’ academic achievements have been improved through CBA.</td>
<td>A  1.3</td>
</tr>
<tr>
<td>25. I am concerned CBA allows more cheating attempts than traditional test.</td>
<td>A  0.4</td>
</tr>
<tr>
<td>26. CBA is more suitable to formative assessment compared to summative assessment.</td>
<td>A -1.4</td>
</tr>
<tr>
<td>27. I don’t think students are complaining about CBA and quizzes given more often.</td>
<td>A -1.0</td>
</tr>
<tr>
<td>28. CBA will be adopted in more learning environments than now with the development of technology.</td>
<td>A  1.2</td>
</tr>
<tr>
<td>29. I find CBA very useful with fast test results and saving, easy accumulation, etc.</td>
<td>A -0.0</td>
</tr>
<tr>
<td>30. The score weights of CBA should be raised (high stakes assessment).</td>
<td>A  0.4</td>
</tr>
<tr>
<td>31. I prefer traditional testing methods.</td>
<td>A -1.3</td>
</tr>
<tr>
<td>32. I think the current level of CBA is enough considering the technical level of CBA.</td>
<td>A -0.7</td>
</tr>
<tr>
<td>33. CBA is most suitable for testing basic knowledge, terminology, and comprehension for skill oriented topics.</td>
<td>A -1.0</td>
</tr>
</tbody>
</table>
Results

The QUANL program was used to analyze the typology of students’ awareness of CBA. As a result of performing QUANL to analyze the Q typology, 30 respondents in the student P sample were divided into four types (refer to Table 3). In this study, the four groups are given the following descriptive labels: (I) CBA dissatisfaction Type, (II) CBA Friendly Type, (III) Adjustment Seeker Type, and (IV) CBA Apprehensive Type. The overall predictable variation of the four types is set to 62.21%; the number of people in each type is: 11 for type I, 6 for type II, 8 for type III, and 5 for type IV. However, no significance should be attached to the number of persons for each type; because the sample size is small, we should not draw any hasty generalizations. This study uses Q methodology only to discern the presence of these types. The factor weight of people who belong to each type indicates the typical person who can serve as a representative of the type that she or he belongs to. For example, this means that #6 (female, sophomore) and #9 (male, freshman) in the P sample are the typical learners (they have factor weights of 3.473 and 3.788 respectively) representing Type I; #23 (male, sophomore) is the person with typical characteristics of Type II.

Next, QUNAL was performed to analyze the typology of professors’ awareness. As a result, P samples of 15 professors have been classified into 3 typologies with factor 1 explaining 25.48% of total variance, factor 2, 20.30%, and factor 3, 6.76%, thus resulting in a total variance explained of 52.54%. Of the 15 respondents, 7 were clustered and interpreted in Type A, 4 in Type B, and 4 in Type C. Factor weight scores for each instructor participant are represented in Table 3. The three groups of instructors’ perception were given the following descriptive labels: (1) CBA Preferred type, (2) Supplementary Need type, (3) and Yes-But Mixed type.

The characteristics of each type can be identified by looking at the average score of agreement or disagreement level for each type among the 42 and 33 Q statements. The aim was to understand the characteristics of each type by extracting a strong agreement (with an average score over +1) and a strong disagreement (with an average score under -1) for statements within the type, in the process of comparing how each type appears for all items (refer to Table 1, 2).

Table 3. Variable assignments with factor weight

<table>
<thead>
<tr>
<th>Participant</th>
<th>Type I (N = 11)</th>
<th>Type II (N = 6)</th>
<th>Type III (N = 8)</th>
<th>Type IV (N = 5)</th>
<th>Factor Weight</th>
<th>Participant</th>
<th>Type I (N = 7)</th>
<th>Type II (N = 4)</th>
<th>Type III (N = 4)</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>.770*</td>
<td>1.890</td>
<td>O</td>
<td>.696*</td>
<td>1.352</td>
<td>15</td>
<td>.711*</td>
<td>A</td>
<td>.662*</td>
<td>1.178</td>
</tr>
<tr>
<td>27</td>
<td>.671*</td>
<td>1.219</td>
<td>J</td>
<td>.621*</td>
<td>1.011</td>
<td>7</td>
<td>.658*</td>
<td>M</td>
<td>.611*</td>
<td>.975</td>
</tr>
<tr>
<td>5</td>
<td>.552*</td>
<td>.794</td>
<td>F</td>
<td>.596*</td>
<td>.924</td>
<td>18</td>
<td>.454*</td>
<td>N</td>
<td>.517*</td>
<td>.705</td>
</tr>
<tr>
<td>19</td>
<td>.358*</td>
<td>.411</td>
<td>D</td>
<td>.427*</td>
<td>.522</td>
<td>23</td>
<td>.864*</td>
<td>L</td>
<td>.811*</td>
<td>2.365</td>
</tr>
<tr>
<td>1</td>
<td>.852*</td>
<td>3.099</td>
<td>K</td>
<td>.698*</td>
<td>1.360</td>
<td>20</td>
<td>.837*</td>
<td>E</td>
<td>.572*</td>
<td>.850</td>
</tr>
<tr>
<td>24</td>
<td>.820*</td>
<td>2.499</td>
<td>B</td>
<td>.461</td>
<td>.585</td>
<td>14</td>
<td>.781*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>.503*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29</td>
<td>.909*</td>
<td></td>
<td></td>
<td>5.221</td>
</tr>
</tbody>
</table>

88
Characteristics of students’ typology

Type I: CBA dissatisfaction type

Type I (N = 11) is named the CBA Dissatisfaction Type. The respondents of this type had a somewhat negative view of CBA. This type tended to prefer face-to-face feedback and paper-based assessment while realizing CBA as a secondary method for student evaluation. They had a skeptical view of students more voluntarily and positively participated in academic learning through CBA; at a time, they considered CBA to be more suitable for basic courses to learn the basic terminology or knowledge. The type and its characteristics of these respondents were clearly revealed in Q statements they strongly agreed with - Q24 (z score = 1.90), Q29 (z = 1.70), Q14 (1.69), Q34 (1.62), Q26 (1.44), and Q30 (1.35), as well as in Q statements they strongly disagreed with - Q31 (z score = -2.20), Q35 (-1.98), Q40 (-1.47), and Q25 (-1.43). The characteristic of Type I is clearer from students’ comments in an interview expressing:

I had a hard time to complete heavy assignment in other subjects. It might be because of my passive personality, I would consider CBA as another assignment given to solve, rather than another opportunity to learn something new. I just completed the CBA assignment formally; I don’t think my autonomy has been improved. (#30, male, freshman)

If I don’t know the answer for a particular problem, I look into the question until I find the answer. But CBA tells the right answer immediately, and this bothers me more. (#6, female, sophomore)

I am used to basic computer activities, such as word processing; however, in my case, a paper-based test helps me to more focus on the test. When I look the question on the computer, I don’t really catch its intention so I write the question down on paper again. I don’t like CBA for math class because I have to rewrite the solving process on the paper. When I learn something new, I print it out, underlin the key points with highlighter, and organize them in my own way. (#27, male, senior)

Type II: CBA friendly type

Type II (N = 6) learners, who are labeled as CBA Friendly Type, expressed positive views about their CBA experiences, which they found stimulating and meaningful. In contrast with Type I, the learners of Type II reported that they have exerted more time and effort on CBA and their academic autonomy has been improved. They are mostly satisfied with the access to CBA and its infrastructure. They found the differences between CBA and paper-based tests, including its provision of graphics, images, and video, to be useful. This Type stated the improved interaction with students and instructors. The respondents in this Type were also satisfied with the support and assistance from their Assistant Instructor. This Type strongly agreed with the Q statements Q22 (z = 2.25), Q17 (2.09), Q21 (1.92), Q11 (1.64), Q31 (1.40), and Q8 (1.39), whereas they strongly disagreed with Q statements Q15 (-

Note. * = Loading factor scores by each type.
1.97), Q35 (-1.74), Q2 (-1.61), Q33 (-1.57), Q3 (-1.36), and Q6 (-1.18) (Refer to table 1). In the in-depth interview additionally conducted, respondents of Type II have confirmed their views on CBA as follows.

“I had an opportunity to perform a very simple CBA in this one class. CBA allowed students to take a test while studying the subject at the same time. This was a great help on our final examination and much more helpful for me to understand the class compared to the traditional classes.” (#14, male, junior)

“Frequent quizzes through the CBA helped students to review and prepare for the classes, much more effectively than general test. I am skeptical of improving academic ability if students cram for one or two-times before big exams.” (#23, male, sophomore)

“My professor uploaded class materials for each chapter in advance to induce students to prepare and pre-study the subject. My professor also gave more practice tests or quizzes. I think this was an opportunity for me to engage in a closer interaction with my professor and other classmates compared to other traditional classes.” (#20, female, sophomore)

**Type III: Adjustment seeker type**

The third perspective \((N = 8)\) is labeled Adjustment Seeker Type. This Type mixed both positive and negative aspects of CBA. This Type recognized the advantages of CBA and at the same time also pointed out the task CBA should solve. As can be seen from the Q statements this type strongly agreed with Q18 \((z = 1.91)\), Q11 \((1.79)\), Q3 \((1.76)\), Q7 \((1.58)\), Q9 \((1.48)\), Q24 \((1.29)\), and Q20 \((1.13)\). The Adjustment Seeker Type appreciates CBA’s flexibility in time and space, improvement of academic comprehension, and immediately feedback test results while pointing out the need for precise feedback and various evaluation methods. This Type strongly disagreed with Q36 \((z = -1.91)\), Q35 \((-1.78)\), Q27 \((-1.52)\), and Q33 \((-1.51)\), indicating the mixed perspectives they have. They considered CBA to be a useful and valuable assessment method and at the same time, they also showed a negative perspective on the test weight to be raised. This is echoed in the following comments from type III respondents:

“Basically I think CBA is more useful in evaluating students that allows me to control my time and place for study and receive the feedback for that. However, the detailed feedback on answers I made incorrect is not enough for some particular subject. In my opinion, it is ineffective for professors to give individual feedback to each student during the class; therefore, the detailed feedback provided by CBA should be improved.” (#29, male, junior)

“I used many other references, internet or textbook, while taking an individual CBA test. It relieves my tension and reduces my effort to study compared to the general test; however, I think CBA should be the supplementary method, not the main method in evaluating students. I think test requiring everyone to come to a particular place at the same time is fairer than CBA” (#17, male, freshman)

**Type IV: CBA apprehensive type**

Type IV \((N = 5)\) is named CBA Apprehensive Type. The respondents of Type IV voiced their strong concerns and worries over heavy tension and cheating attempts. The CBA Apprehensive type had less interest in CBA \((Q13, z = -1.41)\) but a strong desire to get better grades \((Q37, z = -1.00)\). Therefore, they found CBA courses highly stressful throughout the semester because they gave students more assignments to do \((Q4, z = 1.90)\). This type does not trust the system stability of computers in CBA \((Q18, z = 1.52)\) and does not consider CBA a fair assessment either \((Q1, z = -1.97)\), compared to traditional tests, for its high association with cheating attempts \((Q42, z = 1.56)\). The students of this type gave statements as follows.

“2 years ago, I experienced with the vulnerability of servers when everyone in class took a mid-term exam at the same time. A few days ago, I lost my answers in the middle of taking the quiz with my laptop due to my disconnected WIFI.” (#2, female, junior)

“CBA made students attempt cheating in the absence of exam invigilators. Some were sharing their solutions, discussing with others, or taking the exam for others with its simple personal identification procedure. Students always produce a higher average in CBA tests while resulting in under average on paper-based tests. For the math
course, there is an online website that solves mathematical problems instead of the examinee. It is difficult to prevent cheating attempts that are developing from time to time.” (#16, male, sophomore)

Characteristics of instructors’ typology

CBA preferred type

The respondents for instructor’s typology sorted as Type A (N = 7) are labeled CBA Preferred Type. The instructors of this type focused on the benefits of CBA and are satisfied with the CBA they have adopted, showing a positive view of CBA in general. They recognized the advantages of CBA, including the flexibility in class design and operation, utilization of various data, easy grading, improvement of student’s responsibility and academic achievement, and increased interaction between instructors and students. The instructors in this type considered CBA as an interesting and innovative educational tool and responded that they would like to recommend this to other instructors. This type is well revealed in the Q statements with which they have strongly agreed or disagreed: Q2 (z = 1.54), Q16 (1.41), Q14 (1.37), Q24 (1.35), Q17 (1.32), Q8 (z = -2.07), Q15 (-1.44), and Q26 (-1.43). This is also well presented in the following comments.

“I had an opportunity to investigate the status of e-Education at all universities in Korea while preparing e-Education at UNIST. UNIST takes the initiative in Blackboard, such as two-way communication in class with the application of e-Learning contents. Many other universities were also interested in these blended courses and also preparing to adopt them. Therefore, I am sure more innovative and diversified educational tools based on the virtual learning system will be widely used in the learning environment.” (J, Assistant Instructor of Business IT)

“I am combining online and offline classes. I gave my students CBA open-book tests online almost every week. In online classes, I saw my students engaged in individual research activities and self-initiated study by reviewing and looking up new data. My students showed improved academic achievement in the final examination compared to the previous semester without conducting CBA classes.” (O, Professor of Statistics)

Supplementary need type

Type B (N = 4) is named as Supplementary Need Type showed a strong agreement with the following statements: “There are subjects particularly suitable for CBA - Q13 (z = 2.21)”; “Other assessment should be combined in class activities since the question type in CBA is limited - Q11 (z =1.69)”; “Resources to create CBA questions are inadequate - Q21 (z = 1.10)”; “Cheating is a big concern compared to traditional tests - Q25 (z = 1.08)”. On the other hand, the respondents of Type B strongly disagreed with Q statements Q12 (z = -1.62), Q27 (-1.45) and Q3 (-1.11), realizing the need for improved CBA feedback and essay type questions, and greater workload required from students. The instructors of this type were strongly convinced of the need for assistants and administrative support and didn’t positively adopt CBA, in contrast with Type A (Q4, z = -1.25). This Type recognizes current CBA needs to be improved but prefers CBA as a supplementary tool for evaluating students. This Type is well represented in the following comments from the post interview.

“Multilateral, complex evaluation is essential in improving the validity of evaluation; however, CBA is useful for objective tests or participation evaluation. Cheating attempts or technical errors are the issues to be improved. For now, it is good to apply CBA to evaluate students’ preparation for math or physics courses (as basic problems).” (H, Assistant Instructor of Arts)

“It is my idea to selectively, supplementarily apply to some particular subject rather than all subjects.” (N, Professor of Mathematics)

“I could never adopt CBA in the absence of administrative support from the Center of Teaching and Learning, introduction of Blackboard system, assistant instructor’s help in managing online classes or tests. All the support became a great help. However, regardless of their support, I would be better at managing my CBA class if I could utilize the Blackboard system well, including the skills to create, edit, and manage the online questions.” (F, Professor of Philosophy)
Lastly, the third typology of instructor \((N = 4)\) is named the Yes-But Mixed Type. Four instructors loaded significantly onto Type C, showing a positive viewpoint of CBA for improving student’s academic achievement \((Q24, z = 1.02)\), easy tracking of academic activities \((Q19, z = 1.01)\), and satisfactory feedback \((Q12, z = 1.29)\), while presenting the negative perception of CBA for its lack of resources \((Q21, z = 1.08)\) and applicability of subject \((Q13, z = 2.05)\). The inductors in this Type have responded that CBA ultimately needs to be tailored for each students’ characteristic and is more suitable for formative assessment that helps and monitors students, rather than the summative assessment that measures students’ academic achievement. As Type B claimed, they revealed CBA should be a supplementary tool in evaluating students. The following comment in the interview confirms previous claims:

“I use CBA for exercise or attendance check rather than for student evaluation. This CBA is very useful in drawing students into self-initiated study; however, it is not yet suitable for evaluating students properly. It is important to participate in class or team projects for my class. I don’t think CBA should replace the traditional test method.” (K, professor of communication)

**Conclusion and discussion**

The purpose of this study is to provide help in better planning CBA by analyzing the current state of affairs in light of users’ cognition and typology. As a result of this study, four groups for students’ typology were analyzed and given the following descriptive labels: (I) CBA Dissatisfaction Type, (II) CBA Friendly Type, (III) Adjustment Seeker Type, and (IV) CBA Apprehensive Type. Regarding instructors’ typology, three groups were sorted and named: (A) CBA Preferred Type, (B) Supplementary Need Type, and (C) Yes-But Mixed Type. This result shows the diversity of respondent types, from users reluctant to use such learning instrument (CBA) to users preferring and focusing on CBA and its advantages.

Common opinions and characteristics drawn from each type of instructor and student have been organized in <Figure 1> and <Figure 2>. CBA learners mostly agreed with the Q statements, “I prefer more detailed feedback on their answers, especially if they got them wrong (Q20)”, “mechanical approach to learning (Q14)” and “CBA should be a supplementary tool (Q24)” (only the CBA Friendly type feebly agreed). Types I, II and III responded that a lot of time and effort is required for CBA (Q35), whereas types I, III, and IV commonly agreed that increasing the scoring weight of CBA was not needed. Characteristics and common perceptions of each type were also found in the perceptual typology of instructors. All three types commonly recognized the lack of CBA resources (Q21) despite the availability of open education resources (Anderson & McGreal, 2012). Improvement of academic achievement (Q24) and intense workload required from students (Q27) were also commonly shared beliefs among the types. In particular, Type B and C shared a common perception that CBA should be a supplementary tool in evaluating students. If investigating individual characteristics of each type, the difference between each type can be seen as follows in Figure 2 (for example: Type A disagreed with Q8 whereas Type B agreed with it).

The features identifying each of the perceptual types are mostly supported by earlier works, but some distinct features have emerged from this study. For example, the students commonly recognized CBA requires more time and effort than other traditional tests and this fact is in line with the results drawn from the previous study (Liu, 2012). However, a Type IV (CBA Apprehensive) strongly disagreed with the fairness of CBA due to concerns over cheating, showing a completely opposite result to McKenna’s study, which strongly claimed CBA as an extremely fair assessment method for the clarity of its grading scale. In contrast with results claiming student’s improved responsibility for learning through CBA (Marriott and Lau, 2008), more learner types who view CBA as a mandatory and forceful learning tool than an aid to their autonomy or responsibility were found in this typology study, whereas instructors responded their students improved their academic achievement, indicating a result corresponding to the previous study (Zakrzewski & Bull, 1999).

To summarize the present typology study, instructors and students showed mixed viewpoints from welcoming or enjoying to reluctance or dissatisfaction. Each individual type for instructors and students showed distinct characteristics; however, both students and instructors agreed to use CBA as a supplementary tool in evaluating students. Also, both groups commented that CBA needs more time and mental effort than traditional assessments. Studying online learning materials and frequent CBA quizzes or assignment for students and designing and
implementing blended courses for instructors caused high workloads for them. Shared beliefs among student types point out their excessive workload and the need to improve feedback quality, while the instructor types share viewpoints of students’ excessive work just like the students, as well as the lack of CBA resources. Generally, both instructors and students perceived the benefits of CBA and they expressed pitfalls and needs for better application as well. The differentiation between instructors and students types is that instructors tended to seek technical support such as CBA sources, while students were fairly sensitive about the application of CBA. Although there is a CBA friendly type, the shared beliefs among student types shows their apprehensiveness.

Figure 1. Perspectives of the four types on CBA (Students’ typology)

Figure 2. Perspectives of the three types on CBA (Instructors’ typology)
Qualitatively excellent e-feedback has replaced the ineffective CBA feedback nowadays (Yorke, 2003); nevertheless in this study, both instructors and students were convinced of the need to improve CBA feedback. In-depth interviews for this study showed that the negative response toward feedback mostly resulted from insufficient feedback on short or long essay CBA tests. Instructors felt the difficulty in providing meaningful feedback on screen because they faced large numbers of students or practical constraints such as time and workload.

This result may be due to Korea's circumstance and cultural characteristics in education combined with the significance of grades and fierce competition. However, this may become an indicator for the perception and attitude of Blackboard users, for current CBA and blended courses. CBA has reached its particular purposes; however, it is being judged as inadequate for a complete, thorough system. In order to effectively apply CBA to a learning environment, the advantages and disadvantages of CBA, and the perception and attitude of users should be thoroughly investigated, and the purpose of evaluation and its cost-effectiveness should be considered prior to applying CBA.

The result of this study proposes the following technological and pedagogical perspective to faculty, schools, and practitioners. First, the instructors should propose a CBA containing high quality feedback based on students' need in order to maximize its benefits. To do this, the instructors can have access to various CBA resources including questions banks and more technical and administrative support. Second, the result of the present study for investigating the perceptual typology of both instructors and students showed that a complete replacement of CBA or its suitability for summative assessment seemed to take more required time. Instead, a strategy that allows students to move on to the next step is proposed with the application of high quality formative assessment. Third, a more intense academic burden was found in students enrolled in CBA or online courses at a regular university. The instructors should avoid giving too much work to students compared to a traditional class when designing the classes. Lastly, CBA that induces autonomous learning is very suitable for self-regulated learners who rarely depend on instructors (Zimmerman & Schunk, 2001). The present study suggests that provisions or solutions for unmotivated or untalented students are essentially needed.

Acknowledgements

This research was supported by the National Research Foundation in Korea

References


User Acceptance of a Proposed Self-Evaluation and Continuous Assessment System

Antonio Robles-Gómez*, Salvador Ros, Roberto Hernández, Llanos Tobarra, Agustín C. Caminero and José M. Agudo

Control and Communication Systems Department, Spanish University for Distance Education (UNED), Madrid, 28040, Spain // arobles@scc.uned.es // sros@scc.uned.es // roberto@scc.uned.es // llanos@scc.uned.es // accaminero@scc.uned.es // jagudo@gmail.com

*Corresponding author

ABSTRACT
The WWW nowadays enables faculty to develop new Internet-based applications that can be used to enhance classroom instruction. There is a clear evolution towards the implementation of new service-oriented learning/teaching systems, which can be considered as the latest generation of Internet-based platforms. This work first describes a service-oriented system in order to follow the students’ progress and automatically assess their practical activities, not theoretical knowledge. The service oriented paradigm allows us to implement a more advanced system for the self-evaluation and continuous assessment of the students’ progress in the field of Information Technology, where activities measure practical daily-work skills. Second, the Technology Acceptance Model (TAM) is used to analyze the acceptance of our proposal. In the results obtained, the participants found the system significantly easier to use, and it proved useful for their purposes. It is also concluded that participants (students and faculty) feel confident with the system.

Keywords
Distance education, Architectures for educational technology systems, Self-evaluation and continuous assessment, Technology Acceptance Model (TAM), Improving on-line teaching

Introduction

New issues related to the learning process in the field of higher education, especially on-line or distance evaluation, are currently a hot topic of research, and it is of a great interest due to the growing use of Internet (Ryan et al., 2000; O’Reilly, 2007; Dagger et al., 2007; Muñoz-Organero et al., 2010). Basically, the evaluation process consists of determining whether the educational objectives are actually achieved. This fact includes not only the student, but all the elements involved in an educational program (given educational resources, faculty, etc.). Faculty must be able to adapt the learning process to students, strengthening or expanding it as appropriate within the context of the European Higher Education Area (EHEA). In addition to this, if we focus on a distance methodology (Carswell et al., 2000; Morgan et al., 2002; Schrum et al., 2007), the evaluation process becomes an even more important and indispensable instrument within the learning process. By using the available evaluation elements, faculty determines the chosen learning outcomes and, at the same time, the subject content can be adapted dynamically to students. Since interaction between faculty and students within a distance methodology increases during the evaluation process, performing this in an efficient and correct way is essential in order to achieve the previously proposed objectives, as already stated in (Robles-Gómez et al., 2011; Agudo et al., 2011).

This work describes a solution, based on the service-oriented paradigm, to follow the progress of students and, automatically, assess the students’ practical activities. In order to analyze its acceptance, the Technology Acceptance Model (TAM) is adapted for our purposes, since there is a lack of references related to self-evaluation and continuous evaluation service oriented systems. This model is based on the perception of utility and ease of use as key factors when deciding on the use of Information Technologies (ITs). In this sense, TAM questionnaire results for each item will be presented for both students and faculty members.

In order to center this work, the system will only use activities related to the configuration of network services. Its scope is much broader, since this system has been designed and implemented as a modular system, which is independent of the design and implementation of specific activities. In this regard, we focus on the “Network Services Management in Operating Systems” (NetServicesOS) course belonging to the “Communications, Networks, and Content Management” post-graduate program at Spanish University for Distance Education (in Spanish, Universidad Nacional de Educación a Distancia). In our traditional evaluation approach, once a student finished an activity, he/she proceeded to write a solution document detailing how he/she had solved the requested work, and
uploaded it in the virtual platform for its evaluation. After each deadline, faculty could assess the students’ activity from this document, although it was difficult to check the correctness of the activity, since the activities of the subject are very practical which involve starting some network services and, also, creating and modifying a set configuration files, among others (Robles-Gómez et al., 2011; Agudo et al., 2011).

Thanks to the use of the service-based system, faculty can track the progress of a large number of students and adapt dynamically the learning/teaching process. In particular, the system is able to automatically assess the practical activities of the subject, detect possible failures within the network services and configuration files, and help faculty to understand the main weaknesses of students during the process of learning from the evaluation reports that the system generates. These features are not possible in traditional Learning Management Systems (LMSs), since they only focus on multiple choice, questions, short answer questions, or similar, but not practical activities, as our case is. From all this information, faculty will be able to dynamically adapt the content or evaluation resources to the students’ skills if necessary. Students can also receive lively feedback on their activities – which was totally impossible with our traditional evaluation system based on explanation reports for each activity. More details are given in Robles-Gómez et al. (2011) and, afterwards, in Agudo et al. (2011).

Among the principal features of this work, we can find that (1) this paper adapts TAM for educational purposes (2) to study a self-developed automatic evaluation tool from the point of view of both student and faculty participants, (3) which is used in the largest distance University in Spain (over 220,000 students and 2,000 lecturers), (4) in a post-degree subject related to the Computer Science discipline, where practical activities are employed to assess students’ skills, even more, in post-graduate programs. These characteristics jointly make this work different from the literature, which presents results of great interest for researchers and learning practitioners.

This paper is organized as follows. Next section describes the most relevant related approaches found in the literature. After that, this work presents the self-evaluation and continuous assessment service-oriented system. In addition to this, the methodology used to study the acceptance of the system in the field of distance higher education is described and, also, the results obtained from this study are discussed. Finally, the paper outlines our conclusions and possible future work.

Background

Current e-learning platforms mainly belong to the second generation of platforms (Moodle, Sakai, dotLRN…) which can be defined as monolithic (Muñoz-Organero et al., 2010). They provide centralized learning, based on contents distributed by lecturers. These platforms offer evaluation methods focused on theoretical contents, where students can immediately verify their knowledge by multiple-choice tests or exercises of concept associations (Douce et al., 2005). Furthermore, the use of these kinds of technologies provide us with a new set of functionalities, not available in traditional educational environments, such as fast interaction, anywhere and anytime access and usage, and valuable discussions that help the student to improve his/her learning process. In spite of these new capabilities, it is quite common to use traditional evaluation techniques, such as essays or exams, since the costs for the implementation of continuous assessment can be very elevated. However, when preparing educational programs, the execution of a continuous assessment within the students’ evaluation process is really important in order to follow their progress in an efficient way.

According to the implementation of new service-oriented learning/teaching systems, also called third generation platforms (Dagger et al., 2007; Muñoz-Organero et al., 2010), there is a clear evolution towards the usage of personal e-learning environments. State-of-the-art e-learning developments, such as (Wilson et al., 2006; Dabbagh et al., 2012) make use of Personal Learning Environments (PLEs) that are based on services. Service-oriented applications can be considered as the latest generation of Internet-based platforms. They are understood as a set of Web-based services and users are considered as the principal actors of the system. These properties allow faculty to implement more advanced systems, in our case, for the management of the self-evaluation of practical activities and the continuous assessment of the students’ progress. Currently, there are not fully-functional implementations related to these topics jointly, due to the complexity of correction by a computer without having an objective response solution. Faculty chooses non-scalable solutions, often based on proprietary developments to meet part of these needs (Gutiérrez et al., 2010; Lorenzo et al., 2011).
Several efforts have been made towards the automatic evaluation of contents. Guerrero, Bataller, Soria, & Magdalena (2007) presented the BioLab tool which is used for the learning of biosignal processing in the biomedical engineering module of the Electronic Engineering degree at the University of Valencia (Spain). This tool allows results to be obtained interactively during theory lessons and provides support in laboratory sessions. Additionally, Barry (2012) presented a video-recording system for oral presentations of students’ groups, for later viewing and self-assessment by members of the student group. Nevertheless, the main drawback of these two proposals is that, although they provide students with interactive help during some steps of the course, they do not perform an automatic evaluation of the students’ activities. In order to minimize this problem, a specific system for assisting evaluation using test cases was proposed by Lorenzo, Vélez, & Peñas (2011). The main drawbacks of this proposal are that the system only provides a partial-automatic evaluation, based on a set of test cases predefined, it does not implement a continuous assessment process, and authors do not evaluate the students and/or faculty’s acceptance of the system.

On the other hand, Robles-Gómez et al. (2011) outlined the teaching of a course related to the management of network services, which has been adapted to the context of the new EHEA. A first prototype was also defined for the automatic assessment of the students’ activities. In addition to this, Agudo et al. (2011) implemented a particular version, which allow faculty to manage the continuous evaluation of the students’ progress in Linux activities to be used in the subject. For this reason, it has been thoroughly evaluated in this work, by adapting the Technology Acceptance Model (TAM) to our purposes.

**Description of the system**

Service-oriented applications are considered as the latest generation of Internet-based platforms. Using the system, students will be able to perform a self-evaluation of their activities and, additionally, the system will be able to automatically correct a student’s practical activities by modifying their configuration files or, even more, configure it completely, with a penalty in the mark for the activity, as proposed by Robles-Gómez et al. (2011). The system has several main benefits for the members of the learning community, specially, within the field of a distance higher education. First, it minimizes the response time in correcting students’ practical activities, allowing the continuous evaluation process to be performed smoothly. Furthermore, it provides a more detailed monitoring of the students’ progress, thereby reducing the time spent on the assessments themselves. The importance of these benefits is really significant, since the number of students enrolled in a course with a distance methodology can become very high (Robles-Gómez et al., 2011). Thus, faculty can focus on other tasks, such as dynamic adaptation of new activities to students’ necessities or expanding the existing ones, which in turn improves the learning process more than devoting their time to correcting the students’ activities.

The system is made up of two different: the faculty’s view and the students’ view parts (Robles-Gómez et al., 2011; Agudo et al., 2011). From the faculty’s view, faculty can perform subject management tasks such as selecting the activities for the subject, creating different groups with activities adapted to the students’ level, checking students’ progress by means of reports, etc. The Web application, shown in Figure 1, is accessible by faculty through any Internet browser. Note that the system allows faculty to split up the subject’s students by levels or types of activities. In addition, the system provides statistics about the student’s run status and run time. Finally, a list of recent reports is stored for each of them, which can be checked by faculty at will. From the students’ view, the system can automatically configure and/or evaluate a particular activity. For this purpose, the system will download a handler via Internet, and run it inside the students’ infrastructure by using virtual machines, in order to analyze and/or configure the most important configuration elements and generate an evaluation/configuration report. So, students find out which parts of a particular activity are wrong and, additionally, the system can help students when they are not able to do a part of the activity. All this information will be automatically updated on the server side so that it can be used by faculty to improve the learning process.
Architecture of the system

When designing and implementing our service-oriented system, a range of technologies and components were selected, all of them having open source licenses. The system represented in Figure 1 was built using Puppet, which provides most of the relevant components for the automatic evaluation system, and a Ruby on Rails application was developed to meet the needs for the graphical tool. However, we are currently replacing both technologies by only the open-source Flask framework, which is based on Python, in order to have all components of the system, including the self-evaluation pieces and the Web application, fully integrated with the same technology. This will allow us to increase the flexibility of the system, when deploying new services for the system. For both cases, the system follows the Model-View-Control (MVC) paradigm, an architectural design pattern widely used in Internet. The communication between the main elements in the implemented architecture is driven by REpresentational State Transfer (REST) Web-based service interfaces. REST is an architectural software technique for hypermedia distributed systems such as the World Wide Web (WWW). Since it uses HyperText Transfer Protocol (HTTP) as a binding communication protocol, REST services are stateless. This means that each HTTP message contains all the information needed to process a request.

In addition to the REST service properties, in each client/server service-oriented connection, Agudo et al. (2011) proposed a safe SSL channel for encrypting data by means of digital certificates for both endpoints. When starting the request phase, a pair of public/private keys is generated and the certificates needed are signed, automatically or under the supervision of the administrator, by the server as the certificate authority. Security issues will be increased by using the native functionalities provided by the Flask framework. Thus, client impersonation is avoided and there is a suitable security level for remote client configuration. In order to improve the performance and the scalability of our service-oriented evaluation system, given the high number of simultaneous users that this tool must support, Apache HTTP Server and Passenger were chosen for the deployment of our solution. Passenger is a module that mainly helps Apache to improve the performance, scalability, and security issues when deploying the system. Again, when the system is fully-migrated to the Flask technology, our proposed system will achieve the same levels of scalability by simplifying the number of technologies to be employed and, also, by increasing its maintenance.

The implementation of the self-evaluation and continuous assessment system follows a service-oriented paradigm. The concrete architecture of the system is shown in Figure 2, including the interaction among its main elements, which are Web Client, Lab Manager, and Web Server. The Web Client element is executed by the student, at the remote location that he/she uses for the performance of his/her practical activities. It is in charge of connecting with
the Web Server element through the Internet, evaluating the desired configuration (depending on the configured activities at the server) and showing the differences compared with the current configuration. Finally, it must also create an evaluation report with the results of the evaluation and store it in a results database. In addition, the student could run the proposed system so that it automatically configures the student’s local infrastructure. This last operation can only take place with the faculty’s permission, so that the learning and evaluation processes are not jeopardized.

On the other hand, as observed in Figure 2, the Lab Manager allows faculty to assign activities to the students or track their progress in an easy and simple way. It can also provide, on request, the activities associated to students and confirm/modify the grades calculated by the system. It is accessible by faculty through any Internet browser. As a previous step, faculty is in charge of defining a set of practical or real-work activities to be performed by students, in form of manifests. A manifest is the definition of the requirements that must be fulfilled by each activity. Under these considerations, an activity is a group of one or more manifests, such as the configuration of a DNS server, for example, and they are associated to a particular service that must be evaluated. As for the Web Server, it is used to check the students’ configurations. It first establishes the communication between Web Clients and Web Server. Furthermore, the Web Server obtains the activities associated to each student. It also registers the execution of the activities by students. In addition, the Web Server is ready to support more in-depth analysis, so that more knowledge can be inferred from the manifests in order to improve the evaluation process. When the whole system will be migrated to the Flask framework, will be much simpler to analyze all generated information, in order to extract conclusions to be dynamically adapted to the system. Finally, a relational database server, as MySQL is, has been chosen for storing the generated reports, as well as a data model of users, groups, variables and evaluations.

**Methods**

We have conducted a study in order to collect information about usefulness and ease of use of our proposed system, and how comfortable our students are when using it.
Participants

The study conducted was composed of two parts: the first one was focused on students and the second one on lecturers. For the first part of the study, the sample population consisted of 52 students enrolled in the NetServicesOS course belonging to a “Communications, Networks, and Content Management” post-graduate program in the Faculty of Computer Science at our University. In addition, 12 faculty participants were chosen from the Faculty of Computer Science at our University. Neither faculty nor students had previous experience with the system. The idea of working with faculty from computer-related disciplines was to control the variable of computer literacy as a factor in the study, allowing us to focus on their intention to use the system for their computer science subjects, and the trust provided by our proposed system. It has been demonstrated in other works that familiarity with a topic can affect the use of a tool (Belkin et al., 2003).

The profile of the students fits with the standard average of computer science postgraduates. Most of them are Spanish students between 24 and 35 years old with most men (70%) versus women (30%). In this study, we also had two South-American female students. As for faculty, all of them are from Spain between 30 and 50 years old with most men (75%) versus women (25%). For both parts of the study, a qualitative analysis was performed. In this type of analysis, the size of the sample is not a problem since opinions tend to repeat themselves, after a low number of questionnaires (Johnson et al., 2007; De La Fuente-Valentín et al., 2011).

Procedure

The questionnaire was an adaptation of the Technology Acceptance Model (TAM) developed by Davis (1993) to explain how users come to accept and use a technology. This model is based on the Theory of Reasoned Action (TRA) introduced by Fishbein & Ajzen (1975), and it is thought to predict and explain the users’ acceptance of a given technological element (Morris et al., 1997). This model, widely used in different studies related to technology acceptance, is based on the perception of utility and ease of use as key factors when deciding on the use of Information Technologies (IT) in terms of usefulness and ease of use. It has been suggested that the TAM is a cost-effective tool for predicting user acceptance of systems. None of them is related to self-evaluation and continuous evaluation service-oriented systems.

The related works which apply TAM to infer learners’ willingness to accept e-learning systems have also been proposed in recent years (Liaw et al., 2003; Yi et al., 2003; Ong et al., 2004; Pan et al., 2005; Pituch et al., 2006; Ngai et al., 2007; Liaw, 2007; Capra et al., 2007; Raaij et al., 2008; Liu et al., 2010; Curlango-Rosas et al., 2011; Persico et al., 2014). Similarly to these works, TAM will be used to analyze the acceptance to using our service-oriented system, in terms of ease of use, usefulness, and trust, for both student and faculty participants.

Table 1. Adapted TAM questionnaire statements used to measure students’ perceptions in terms of ease of use and usefulness, and additional statements related to students’ trust

<table>
<thead>
<tr>
<th>Student’s statements (SS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS1  Learning to use the SYSTEM to configure and check my activities is easy for me.</td>
</tr>
<tr>
<td>SS2  I would find it easy to get the SYSTEM to do what I want it.</td>
</tr>
<tr>
<td>SS3  My interaction with the SYSTEM is clear and understandable.</td>
</tr>
<tr>
<td>SS4  I find the SYSTEM easy to use to configure and check my activities.</td>
</tr>
<tr>
<td>SS5  I find the SYSTEM flexible to interact with.</td>
</tr>
<tr>
<td>SS6  It is easy for me to become skillful at using the SYSTEM to configure and check my activities.</td>
</tr>
<tr>
<td>SS7  Using the SYSTEM to configure and check my activities can enable me to accomplish my work more quickly.</td>
</tr>
<tr>
<td>SS8  Using the SYSTEM to configure and check my activities can improve my learning.</td>
</tr>
<tr>
<td>SS9  Using the SYSTEM to configure and check my activities in my job can improve my productivity.</td>
</tr>
<tr>
<td>SS10 Using the SYSTEM to configure and check my activities can enhance my effectiveness.</td>
</tr>
<tr>
<td>SS11 Using the SYSTEM to configure and check my activities can make it easier to do my work.</td>
</tr>
<tr>
<td>SS12 I find the SYSTEM useful to perform my activities.</td>
</tr>
<tr>
<td>SS13 My trust in the SYSTEM is enough when configuring and checking my activities.</td>
</tr>
<tr>
<td>SS14 My trust in the SYSTEM allows me to improve my learning.</td>
</tr>
<tr>
<td>SS15 The trust that the SYSTEM provides me with is satisfactory.</td>
</tr>
</tbody>
</table>
Therefore, participants were asked to express how useful and easy to use they considered the system in order to automatically evaluate their progress and, where necessary, configure some parts of their practical assignments. Table 1 shows the adapted TAM statements from the point of view of students. For this study, the word SYSTEM was shown to participants. Participants were presented with the statements in Table 1 in random order and asked to indicate how they felt about each statement. Statements 1-6 (SS1-SS6) measure participants’ perceptions of the system’s ease of use. Statements 7-12 (SS7-S12) measure participants’ perceptions of the system’s usefulness. Similarly, additional statements (SS13-SS15) were asked to measure participants’ perceptions of the system’s trust. The items used in the survey have been adapted from the previous works (Davis, 1993; Curlango-Rosas et al., 2011). Each item is a five-point Likert-type scale ranging from (1) “Strongly Disagree” to (5) “Strongly Agree” (Davis, 1993; Capra et al., 2007). A similar procedure was followed for the second part of the study, in which faculty participants were asked to follow the status of the students’ activities, examine the functionality of the system, and proceed to check the correction of the activities implemented by students throughout the duration of the NetServicesOS course. These statements are shown in Table 2.

Table 2. Adapted TAM questionnaire statements used to measure faculty’s perceptions in terms of ease of use and usefulness, and additional statements related to faculty’s trust

<table>
<thead>
<tr>
<th>Faculty’s statements (FS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS1 Learning to use the SYSTEM to manage students’ progress is easy for me.</td>
</tr>
<tr>
<td>FS2 I would find it easy to get the SYSTEM to do what I want it to do.</td>
</tr>
<tr>
<td>FS3 My interaction with the SYSTEM is clear and understandable.</td>
</tr>
<tr>
<td>FS4 I find the SYSTEM easy to use to manage students’ progress.</td>
</tr>
<tr>
<td>FS5 I find the SYSTEM flexible to interact with.</td>
</tr>
<tr>
<td>FS6 It is easy for me to analyze students’ skillfulness at using the SYSTEM.</td>
</tr>
<tr>
<td>FS7 Using the SYSTEM to manage a student’s progress can enable me to accomplish my job more quickly.</td>
</tr>
<tr>
<td>FS8 Using the SYSTEM to manage a student’s progress can improve my teaching.</td>
</tr>
<tr>
<td>FS9 Using the SYSTEM to manage a student’s progress can improve my productivity.</td>
</tr>
<tr>
<td>FS10 Using the SYSTEM to manage a student’s progress can enhance my effectiveness.</td>
</tr>
<tr>
<td>FS11 Using the SYSTEM to manage a student’s progress can make my job easier to do.</td>
</tr>
<tr>
<td>FS12 I find the SYSTEM useful to perform my job.</td>
</tr>
<tr>
<td>FS13 My trust in the SYSTEM is enough when tracking my students’ progress.</td>
</tr>
<tr>
<td>FS14 My trust in the SYSTEM allows me to use it during the evaluation process.</td>
</tr>
<tr>
<td>FS15 The trust that the SYSTEM provides me with is satisfactory.</td>
</tr>
</tbody>
</table>

Results

In this section, we present results from both parts of our study. The TAM questionnaire results for each item are presented for all participants (students and faculty). In particular, we show median and mean values, and standard deviation for all collected data. After that, the global user acceptance scores are summarized for each construct (ease of use, usefulness, and trust) in order to compare student and faculty results.

Student and faculty TAM questionnaire results

We obtained the responses to all statements in the TAM section of the questionnaires. Table 3 shows the median and mean values, as well as standard deviation of the responses for the statements regarding participants’ perceptions (students and faculty) of our proposed system’s ease of use (SS1-SS6 and FS1-FS6), usefulness (SS7-S12 and FS7-FS12), and trust (SS13-SS15 and FS13-FS15). All values shown range from (1) “Strongly Disagree” to (5) “Strongly Agree.”

First, we explain the data collected from the student survey, which corresponds to the mean values of the left side of Table 3. It can be observed that learning to use the proposed system when configuring and checking activities is really easy for the students enrolled in the subject (SS1 – 4.07). For this statement, its median value corresponds to 4, which can be considered as “Really Agree.” For the next statement, students agree it is easy to get the system to do what they want (SS2 – 3.66), and they do not agree or disagree that their interaction with the system is clear and understandable (SS3 – 2.98). In these cases, median values correspond to 3.5 and 3, respectively, which can be
considered as “Agree” and “Neither Agree nor Disagree” for each of them. Both of them may be the weakest values in the survey, since their interaction with the system was rudimentary, so it is not intuitive at first. In contrast to this, and a really positive point, students find the system easy to use in order to configure and check their activities, and they also find it flexible to interact with (SS4 – 4.11 / SS5 – 3.75, respectively). In general, it is easy for students to become skillful at using the system to configure and check their activities (SS4 – 3.74).

According to the usefulness of the system for managing students’ progress, faculty participants agree the system makes them accomplish their work more quickly, since the median value obtained is 4 (SS7 – 3.83). In addition, they fully agree that the system is really able to improve their learning and suitable when configuring and checking their activities (SS8 – 4.66). Also, students agree the system improves their productivity (SS9 – 3.73). Additionally, students fully agree that the system enhances their effectiveness and makes it easier to do their work (SS10 – 4.43 / SS11 – 4.47, respectively). In general, it can be observed that students agree about the usefulness of the system in order to perform their activities (SS12 – 4.77). Note that the median value of this last statement is 5, which corresponds to “Strongly Agree.” As for students’ trust in the system, they find the system totally trustworthy (SS15 – 4.43), since its median value is 4.5. More in detail, many students agree that the trust provided by the system is enough when checking their activities (SS13 – 3.79). They fully agree that their trust in the system allows them to improve their learning (SS14 – 4.25).

We now describe the data collected from the faculty survey, which corresponds to the mean values of the right side of Table 3. It can first be observed that faculty participants fully agree that learning to use the proposed system in order to manage students’ progress is easy (FS1 – 4.45), since its median value is 4.5. Likewise, faculty participants really agree that it is easy to get the system to do what they want (FS2 – 4.11). In addition, most of them agree that their interaction with the system is clear and understandable (FS3 – 3.73). The reason could be that they interact with the system by means of an advanced graphical interface. Furthermore, they find the system really easy to use in order to manage students’ progress (FS4 – 4.34), and they find it flexible to interact with (FS5 – 3.91). This last value is slightly lower, since faculty participants are more conscious this type of systems could be vulnerable. This is not our case thanks to the encrypted communication included within the system. Faculty participants agree the system makes it easy to analyze students’ skillfulness (FS6 – 4.47), since its median value is 4.5.

According to the usefulness of the system for managing students’ progress, faculty participants really agree that the system enables them to accomplish their job more quickly (FS7 – 4.00). They fully agree that the system is able to improve their teaching and productivity (FS8 – 4.56 / FS9 – 4.40, respectively), since their median values are 4.5. Also, most of the faculty participants strongly agree that the system enhances their effectiveness and, similarly, on

<table>
<thead>
<tr>
<th>Construct</th>
<th>Students’ statement</th>
<th>Median (mean)</th>
<th>Stand. dev.</th>
<th>Faculty’s statement</th>
<th>Median (mean)</th>
<th>Stand. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of use</td>
<td>SS1</td>
<td>4 (4.07)</td>
<td>0.90</td>
<td>FS1</td>
<td>4.5 (4.47)</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>SS2</td>
<td>3.5 (3.66)</td>
<td>1.05</td>
<td>FS2</td>
<td>4 (4.11)</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>SS3</td>
<td>3 (2.98)</td>
<td>1.23</td>
<td>FS3</td>
<td>3.5 (3.73)</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>SS4</td>
<td>4 (4.11)</td>
<td>0.88</td>
<td>FS4</td>
<td>4.5 (4.34)</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>SS5</td>
<td>4 (3.75)</td>
<td>1.12</td>
<td>FS5</td>
<td>4.0 (3.91)</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>SS6</td>
<td>3.5 (3.74)</td>
<td>0.93</td>
<td>FS6</td>
<td>4.5 (4.47)</td>
<td>0.67</td>
</tr>
<tr>
<td>Usefulness</td>
<td>SS7</td>
<td>4 (3.83)</td>
<td>1.1</td>
<td>FS7</td>
<td>4 (4.00)</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>SS8</td>
<td>4.5 (4.66)</td>
<td>0.53</td>
<td>FS8</td>
<td>4.5 (4.56)</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>SS9</td>
<td>3.5 (3.73)</td>
<td>1.03</td>
<td>FS9</td>
<td>4.5 (4.40)</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>SS10</td>
<td>4.5 (4.43)</td>
<td>0.57</td>
<td>FS10</td>
<td>4.5 (4.47)</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>SS11</td>
<td>4.5 (4.47)</td>
<td>0.64</td>
<td>FS11</td>
<td>4.5 (4.47)</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>SS12</td>
<td>5 (4.77)</td>
<td>0.40</td>
<td>FS12</td>
<td>4.5 (4.71)</td>
<td>0.45</td>
</tr>
<tr>
<td>Trust</td>
<td>SS13</td>
<td>4 (3.79)</td>
<td>1.05</td>
<td>FS13</td>
<td>4 (3.91)</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>SS14</td>
<td>4.5 (4.25)</td>
<td>0.62</td>
<td>FS14</td>
<td>4 (3.93)</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>SS15</td>
<td>4.5 (4.43)</td>
<td>0.64</td>
<td>FS15</td>
<td>4 (4.24)</td>
<td>0.65</td>
</tr>
</tbody>
</table>
the fact that the system makes it easier to do their job (FS10 and FS11 – 4.47). In general, it can be observed that faculty participants find the system suitable when performing their job (FS12 – 4.71).

As for faculty’s trust in the system, they find the system really trustworthy (FS15 – 4.24), since its median value is 4. More in detail, faculty participants really agree that the trust provided by the system is enough when tracking students’ progress (FS13 – 3.91). They also agree that their trust in the system allows them to use it during the evaluation process (FS14 – 3.93). Note that the standard deviation of all presented values is quite low. This fact implies that many participants have similar opinions with regarding most of the statements. To finish this section, it can be concluded from the previous explanations that the proposed system is easy to use, although students have a weaker opinion than faculty. Both of them really agree about the usefulness of the system, and their trust is very high.

User perceptions using percentages

In this section, results for the participants’ perceptions are detailed using percentage values. Table 4 presents these results for the students’ perceptions, analyzing their ease of use, usefulness, and trust. In this case, percentages are shown for all statements with respect to each possible response. It can be observed that learning to use the proposed system when configuring and checking activities is really easy for students enrolled in the subject, since 92% of them agree on this (those who answered 4 or 5 to SS1). Also, 80% of students agree it is easy to get the system to do what they want (those who answered 4 or 5 to SS2). In contrast, only 58% of students think their interaction with the system is clear and understandable (see results from SS3). The reason could be that their interaction with the system was still rudimentary, but it is most important that they do find the system easy to use to configure and check their activities and flexible to interact with (see results from answers 4 or 5 to SS4 and SS5, respectively). Note that, in general, it is easy for students to become skillful at using the system to configure and check their activities, since 84% of them agree on this.

According to the usefulness of the system for the configuration and checking the correctness of activities, 74% of students think it enables them to accomplish their work more quickly (those who answered 4 or 5 to SS7). In addition they believe the system is really able to improve their learning and productivity when configuring and checking their activities, since 96% and 84% of students agree on this, respectively (see results from SS8 and SS9). Furthermore, students may well think that the system enhances their effectiveness because of the fact 96% of them agree on this and, similarly, 92% of them agree on the fact that the system makes it easier to do their work (those who answered 4 or 5 to SS10 and SS11). It can be observed that 100% of students find the system suitable when performing their activities (see results obtained from SS12). As for students’ trust in the system, in general, the trust it provides is satisfactory, since 92% of students agree on this (see results from SS15). More in detail, 84% of students’ trust is
enough when checking their activities (those who answered 4 or 5 to SS13). Also, 92% of them believe that their trust in the system allows them to improve their learning (see results from SS14).

As for faculty, Table 5 presents the results obtained for the faculty’s perceptions using percentage values, in order to analyze ease of use, usefulness, and trust. The conclusions obtained from these results, not detailed due to the lack of space, are more or less similar to the students’ perceptions.

Table 5. Results for faculty’ perceptions in terms of ease of use, usefulness, and trust (%)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Faculty’s statements (FS)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of Use</td>
<td>FS1</td>
<td>0 %</td>
<td>0 %</td>
<td>8 %</td>
<td>25 %</td>
<td>67 %</td>
</tr>
<tr>
<td></td>
<td>FS2</td>
<td>0 %</td>
<td>0 %</td>
<td>16 %</td>
<td>42 %</td>
<td>42 %</td>
</tr>
<tr>
<td></td>
<td>FS3</td>
<td>0 %</td>
<td>8 %</td>
<td>17 %</td>
<td>42 %</td>
<td>33 %</td>
</tr>
<tr>
<td></td>
<td>FS4</td>
<td>0 %</td>
<td>0 %</td>
<td>17 %</td>
<td>17 %</td>
<td>66 %</td>
</tr>
<tr>
<td></td>
<td>FS5</td>
<td>0 %</td>
<td>8 %</td>
<td>17 %</td>
<td>17 %</td>
<td>58 %</td>
</tr>
<tr>
<td></td>
<td>FS6</td>
<td>0 %</td>
<td>0 %</td>
<td>8 %</td>
<td>25 %</td>
<td>67 %</td>
</tr>
<tr>
<td>Usefulness</td>
<td>FS7</td>
<td>0 %</td>
<td>0 %</td>
<td>25 %</td>
<td>33 %</td>
<td>42 %</td>
</tr>
<tr>
<td></td>
<td>FS8</td>
<td>0 %</td>
<td>0 %</td>
<td>8 %</td>
<td>17 %</td>
<td>75 %</td>
</tr>
<tr>
<td></td>
<td>FS9</td>
<td>0 %</td>
<td>0 %</td>
<td>8 %</td>
<td>33 %</td>
<td>59 %</td>
</tr>
<tr>
<td></td>
<td>FS10</td>
<td>0 %</td>
<td>0 %</td>
<td>8 %</td>
<td>25 %</td>
<td>67 %</td>
</tr>
<tr>
<td></td>
<td>FS11</td>
<td>0 %</td>
<td>0 %</td>
<td>8 %</td>
<td>25 %</td>
<td>67 %</td>
</tr>
<tr>
<td></td>
<td>FS12</td>
<td>0 %</td>
<td>0 %</td>
<td>8 %</td>
<td>25 %</td>
<td>75 %</td>
</tr>
<tr>
<td>Trust</td>
<td>FS13</td>
<td>0 %</td>
<td>0 %</td>
<td>17 %</td>
<td>66 %</td>
<td>17 %</td>
</tr>
<tr>
<td></td>
<td>FS14</td>
<td>0 %</td>
<td>0 %</td>
<td>25 %</td>
<td>42 %</td>
<td>33 %</td>
</tr>
<tr>
<td></td>
<td>FS15</td>
<td>0 %</td>
<td>0 %</td>
<td>8 %</td>
<td>50 %</td>
<td>42 %</td>
</tr>
</tbody>
</table>

From the results shown in this section, as in previous sections, we can also conclude that perceptions (in terms of ease of use, usefulness, and trust) are really positive for both the student and faculty surveys. Consequently, the proposed system had a great encouraging impact on the educational process from the point of view of both the student and faculty participants.

Conclusions

The use of Internet-based platforms to allow faculty to enhance classroom instruction is a field with a series of challenges to be overcome. In our particular case, the search for new solutions for developing e-learning systems to follow the progress of students and automatically assess the students’ practical activities is essential within a higher educational system, this being especially true in the case of the EHEA. As a consequence, the main contributions of this work are the following. Firstly, we describe a service-oriented system to manage the self-evaluation and continuous assessment of the students’ progress in the field of Information Technology, where activities measure practical daily-life skills, and some guidelines of how we are evolving the system in order to increase its general flexibility from our several years’ experience. Therefore, the solutions of the associated practical activities to the service-oriented system are subjective. From our knowledge, there was not existing implementations to perform self-evaluation and continuous assessment of these types of activities. Secondly, this fact is even more important when using a distance methodology, as is the case of our University, since the number of students enrolled for each course can become elevated. The most immediate consequence is that the response time in checking the students’ practical activities is minimized. In addition, automatic evaluation of practical activities allows faculty to reduce the time spent on this process and, thus, increase the time available for the dynamic adaptation of the learning process to students. They are also free to evaluate their progress in an immediate and reliable way, so they can independently go deeper in carrying out the practical activities, and it is not necessary to wait for their final evaluation. Thirdly, to conduct our research on the students’ perceptions, we have chosen the NetServicesOS course, where students are expected to deploy and configure network services (DNS, DHCP, Web, etc.). Finally, the Technology Acceptance Model (TAM) is adapted to analyze the acceptance to use the system, in terms of ease of use, usefulness, and trust, for both student and faculty participants.
Within the scope of the activities that have been covered throughout the duration of this course, it has been particularly complex to develop the system due to the diversity of mechanisms involved in it (configuration files, software bases or implementation units) and the speed of software and operating system evolution. Therefore, we have implemented a solution that is as generic as possible, ignoring solutions that would tie us to using a specific software or version (that is difficult to change), which would not allow the system to evolve, as we have described above. The automatic evaluation part of the system allows us to define the requirements to be satisfied for each of the activities in a modular and expandable way. In order to adapt the activities to the system it is necessary to make a statement as specific as possible of those real-work practices, establish the evaluation elements, and the definition of them including the configuration files and templates to be used during the students’ continuous assessment process. In addition, users of the system, the students in the initial scope of its application, are free to evaluate their progress in an immediate and reliable way, so they can independently go deeper in carrying out the practical activities. This fact enables the evaluation process to be automated and transferred to the students, and it is not necessary to wait for their final evaluation.

According to the data collected from the surveys, the student and faculty participants who evaluated the system found it significantly easier to use and useful for the purpose of dynamically following the students’ progress and automatically evaluating their practical activities. Therefore, the system had a great positive impact on the educational process from the point of view of the student and faculty participants. It was also concluded that participants were confident with the system. In this sense, different results were described and more detailed conclusions were given in the corresponding sections for each statement. Finally, some participants suggested possible improvements for our proposed system, most of them have already been incorporated into it. In addition to this, the fact that our study only focused on users from one particular field of study, namely computer science, is a limitation. A possible future work would be focused on working with participants from other fields of study.

Acknowledgements

Authors would like to acknowledge the support of the following European Union project: MUREE (530332-TEMPUS-1-2012-1-JO-TEMPUS-JPCR); and a Local Project (2014-027-UNED-PROY) from the Spanish University for Distance Education. Furthermore, we thank the Region of Madrid for the support of E-Madrid Network of Excellence (S2013-ICE2715). Authors also gratefully acknowledge participants that took part in the surveys.

References


Liu, I.F., Chen, M. C., Sun, Y. S., Wible, D., & Kuo, C. (2010). Extending the TAM model to explore the factors that affect intention to use an online learning community. *Computers & education, 54*(2), 600–610.


Group Learning Assessment: Developing a Theory-Informed Analytics

Wanli Xing, Robert Wadholm, Eva Petakovic and Sean Goggins*
University of Missouri, Columbia, MO, 65203, USA // wxdg5@mail.missouri.edu // rrwb84@mail.missouri.edu // ehpppf@mail.missouri.edu // GogginsS@missouri.edu

*Corresponding author

ABSTRACT

Assessment in Computer Supported Collaborative Learning (CSCL) is an implicit issue, and most assessments are summative in nature. Process-oriented methods of assessment can vary significantly in their indicators and typically only partially address the complexity of group learning. Moreover, the majority of these assessment methods require time-intensive coding of qualitative data. Our study explores the operationalization of activity theory to frame group activity in a CSCL context by breaking group work into six dimensions. We map log data generated by a collaborative software, Virtual Math Teams with GeoGebra, with these dimensions and construct six measures to lay the groundwork for automating CSCL assessment. Next, we move beyond identification and analysis of those measures to infer group learning using human judgment and employ a clustering algorithm to categorize groups with similar performance, allowing us to consider the six indicators simultaneously and to step further toward assessment automation. Last, in terms of the complexity of group learning in a socio-technical context, we discuss a web-based tool that not only shows group-level assessment but also integrates with our previous work on individual assessment, thus providing teachers with a diverse of group learning.

Keywords

Assessment, CSCL, Activity theory, Cluster analysis, Automation

Introduction

A growing body of research has coalesced around the idea that groups, not individuals, are the principle engines of learning (Strijbos, 2011; Stahl, 2006). According to Vygotsky (1978), high-level cognitive functions appear first as interspsychological processes, or group learning, and only later as intrapsychological learning, which results from the internalization of social participation. Learning in groups is a matter of participation in social processes and interactions with artifacts (Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003). According to Stahl (2006), group cognition presupposes three levels of learning: individuals, small groups, and communities. Different analyses and assessment methods may occur at each level, and these levels influence each other and are best viewed as an integrated, complex whole: understanding a group in its context becomes a key to measuring learning individually and at the community level.

However, assessment of in-group learning remains by-and-large summative in nature (Gress, Fior, Hadwin, & Winne, 2010): the qualities of group outcomes or collaboration products are considered the key criterion for assessing collaborative learning. While “real-time,” “process-orientation” and “socio-technical context” are key characteristics of the group concept in collaborative learning (Sfard, 1998; Reimann, 2009), summative evaluations are usually administered after the collaboration, which fundamentally undermines the theoretical constructs behind group learning in CSCL. Assessments conducted during collaboration have various constructs and interests and usually address part of group dynamics and collaboration for learning. On the other hand, a majority of assessment methods during collaboration have consisted of observations and content and interaction analyses (Strijbos, 2011; Gress et al., 2010). Coding these kinds of data is usually time-intensive (Daradoumis et al., 2006). It is difficult, if not impossible, to implement this kind of group assessment on a large scale.

Our study explores operationalizing activity theory to frame group activity in a CSCL context by breaking down group work into six dimensions. Then, rather than performing observations or content analysis to generate our measures, we construct measures based on electronic trace data generated by collaborative software. Next, we move beyond identification and analysis of those measures to infer group learning using human judgment, but employ a clustering algorithm to categorize groups with similar participation performance and further automate assessment. Last, in terms of the complexity of collaborative learning in a socio-technical context, we discuss how a web-based tool that is in development integrates with our previous work on individual assessment, providing teachers with a holistic and multi-perspective view of group learning.
Literature review

The activity of assessment can strongly influence learning (Gress et al., 2010). Research on collaborative learning assessment has largely followed two lines: assessing the individual and assessing the group. Although studies diverge along these two lines, Sfard (1998) proposes a theoretical unification of collaborative learning and assessment, arguing that to focus only on individuals or groups may lead to distortions and undesirable practices. Xing, Wadholm and Goggins (2014a) conducted an extensive review of assessment methods and analysis at the individual level. The authors constructed an activity theoretic, quantified model consisting of six measures to holistically gauge individual participation in group learning activities based on electronic trace data. Teachers can then use these constructed measures to identify and provide feedback regarding the shortcomings of students.

Assessment of groups in collaborative learning has been dominated by “after collaboration” measurement (Gress et al., 2010). After reviewing 186 articles, Gress et al. state that over half of studies measuring collaborative learning were conducted after collaboration. Typically, groups in CSCL are evaluated through the assessment of final products. Kapur and Kinzer (2009) looked at productive failure in CSCL by comparing groups solving ill- and well-structured problems. The performance of each group was measured by the quality of the solution produced, using a rubric for assessment.

Assessments performed during collaboration have a variety of foci. Gress et al. (2010) found 71 instances of metrics gauging group collaboration in four categories derived from prior literature: first, the investigation of successful collaboration. For instance, Safin et al. (2010) video-recorded the collaborative activities of two teams in the domain of architectural design, assessing for successful collaboration based on their own protocol developed from seven dimensions including fluidity of collaboration, task and time management, and individual task orientation. Arnold, Ducate, Lomicka, and Lord (2009) studied foreign language graduate student collaboration in a wiki context. Through content analysis of the dialogue with reference to Curtis and Lawsons’ Framework, they constructed five measures to determine successful student collaboration.

The second construct focuses on group performance using GeoGebra. This builds on the work of Fessakis, Dimitracopoulou, and PalaioDIMONOS (2013), who developed a graphical interaction analysis tool to investigate activity levels within and among groups in a blog environment by counting connections with other members and number of posts. Their results show a significant impact of the graph’s presence on group interaction. The interaction graph generation process used for assessment was semi-automated in this research. Our third research construct focuses on social interactions and communication within collaborative settings. We build on Calvani, Fini, Molino, and Ranieri (2010), who proposed eleven indicators to measure the effect of interaction. Their group indicators relied on content analysis of group dialogue and on message counts.

The last area of focus of group assessment is knowledge construction and skill acquisition. Kumar, Gress, Hadwin, and Winne (2010) built a tool to inspect group knowledge construction and cognition using an ontological approach. They tried to consider the full range of collaborative activities that relate to group learning. The central question is “How does one go about designing a system that affords multiple collaborative opportunities while capturing all learner activities and the full context of their learning environments (e.g., open documents, notes being made, searches, conversations, and sharing) all on a standardized underlying metric…?” They built their own ontologies to assess those aspects based on electronic trace data. Milrad (2002) employed construction kits, modeling tools, and system dynamics simulations to inquire into collaborative discovery learning. All collaborative activities were recorded and transcribed to assess students’ skills and knowledge acquisition.

In summary, group assessment in CSCL mainly takes place after collaboration and focuses on the collaborative product rather than the process. Due to the diversity of process-oriented measures for group assessment in the CSCL community, research explorations vary widely with different interests and constructs covering artifacts, context, interactions, and knowledge development. However, few studies are able to evaluate groups in CSCL holistically. In addition, assessments with process-related metrics rely too heavily on conventional methods such as content analysis, coding of observations and video recordings. These measures provide invaluable information, but since this analysis is very time-intensive and is conducted post-event, information obtained does not offer opportunities for real-time feedback to enhance evaluation, reflection, awareness, and adaptation (Kumar et al., 2010). Though there are different perspectives on group learning, the most common ones share a focus on participation as a condition for learning. Our method approaches group assessment from an activity theory perspective and centers on participation.
Theoretical framework

The Activity System model developed by Engestrom (1987) offers a way to comprehensively frame the collaborative knowledge development process while linking together social behavior and its interdependencies. To illustrate, an activity system provides three characteristics for analyzing learning in group work (De Laat, 2006): activity theory focuses on contextuality and is oriented towards systematically comprehending group dynamics, objects, mediating artifacts and social organization; it also relies on a dialogical theory of knowledge and thinking (language and communication) with a concentration on human cognition; further it is a developmental theory aiming to explain changes in human practices over time.

Engestrom’s (1987) activity model includes six interacting components: Subjects, Rules, Tools, Community, Division of Labor, and Object (see Fig. 1). The activity of learning is “the joint activity of a student, physical/symbolic tool(s), and another person(s) performing together as a working social system to achieve some outcome under constraints such as rules.” In our CSCL group assessment context, the process and outcome of this transformation may both be seen as learning and knowledge. It is the sum of the system components and the tensions among them that make up learning and influence learning outcomes.

Current assessment practices of group performance often address only part of the activity of the learning system, focusing on process, context, cognition, artifacts or a mix of two or three such components. Activity theory allows us to address the complex interactions and envision group performance in a socio-technical CSCL environment (see Table 1).

**Table 1. Description of activity theory operationalization**

<table>
<thead>
<tr>
<th>Measure-metric</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Individual student effort that contributes to problem solving.</td>
</tr>
<tr>
<td>Rules</td>
<td>Implicit and explicit rules and guidelines constraining the activity. For example, teachers can set specific rules for a learning task (explicit), and students can use only the functions supported by the available tools and are bounded by socially established (implicit) norms.</td>
</tr>
<tr>
<td>Tools</td>
<td>Computers, online tools, systems, and environments that mediate the learning and collaboration activities.</td>
</tr>
<tr>
<td>Community</td>
<td>Direct and indirect communication enabling the group of students to maintain a sense of community and belonging</td>
</tr>
<tr>
<td>Division of labor</td>
<td>Overall coordination among group members.</td>
</tr>
<tr>
<td>Object</td>
<td>Jointly completed learning tasks such as the solution of a problem or production of an artifact (e.g., essays).</td>
</tr>
</tbody>
</table>
Research context

VMT

In this study, we operationalize activity theory in order to make sense of electronic trace data to assess groups using a synchronous math discourse tool. We focus on several modules of a 2013 course designed for Virtual Math Teams with Geogebra (VMTwG) software (Figure 2).

The five modules analyzed include teams with 3-5 members who were going through the curriculum using VMTwG. Fourteen groups were chosen for this study. The 5 modules that were analyzed included: “Constructing Dynamic-Geometry Objects,” “Exploring Triangles,” “Creating Construction Tools,” “Constructing Triangles,” and “Inscribing Polygons.” The full curriculum currently includes a total of 21 topics, and is available at the project website (http://vmt.mathforum.org).

Figure 2 provides a guide to understanding cognitive learning discourse in VMT. Section A of Figure 2, the VMT replayer bar, represents the time dimension. Each action within VMTwG is logged with a timestamp. Section B is the chat window, where text is entered into chat. Future analytics in this project will focus on the analysis of the text in chat windows, in concert with GeoGebra gestures. Sections C and D are related to GeoGebra actions. C is the “Take Control” button, which gives an individual user control of the tools. Section D is the GeoGebra window. Here, students use multiple approaches to inscribe an equilateral triangle within another equilateral triangle.

Figure 2. VMTwG, an analytical tool for collaborative math discourse

Dataset description

All log data for this study was collected in .txt format, which centers on specific event types: VMT Awareness, GeoGebra, System, Chat, and WhiteBoard (Wb). The Chat event type logs all the messages in the group. Awareness records the action of erasing chat messages. GeoGebra logs information on how students virtually construct a geometry artifact (adding a point, updating a segment). The system logs all user actions, and automates the construction of action adjacencies to reveal the order each participant works in. Figure 3 shows a sample of original log data.
Methodology

Data categorization

Our analysis examines each interaction as related to four dimensions for a single student in each group: Individual, Group, Action Types, and Module. The Individual category sums all the personal actions in every event type for each individual member of the group (frequency in each event type) in the situation where the student is both the creator (source) and the receiver (target) of the action according to Figure 3. Similarly, the Group category sums all actions taking place between a student and their team members (where the student is the creator and the team member is the receiver). The Action Types dimension is a set of values containing the kinds of actions over that particular event type. For example, if a student never erased a message in the Awareness event over all the modules, then the action types for Awareness is {0}; for a Wb event, if a student takes actions such as creating a textbox and copying an object, but never uses other actions such as moving objects, resizing, etc., then the action types in Wb are {create a textbox, copy an object}; for the System event, a student’s action constraints might be represented as {Join a room, View a Tab}. The awareness dimension is noted according to curricular module: {Module 1, Module 3, Module 4}.

Data formation

Log data is then processed to center on the group as a unit. \( m \) notates the group, \( m=1,2,\ldots,M \). In this context, \( M \) is 14 indicating that there are 14 groups in total. \( N \) represents the number of students in that particular Group \( m \), \( n=1,2,\ldots,N \). \( N \) may be different for different groups. \( a \) represents a single student. \( T \) denotes the event type, \( T_0 \) represents the value of measurement aspects \( j \) in event type \( i \), where \( i \in [1,5] \), \( j \in [1,4] \), \( i \in Z^+ \), \( j \in Z^+ \). To be more specific, there are five event types (i.e., Awareness, Geogebra, System, Chat, and Wb) and four measurement aspects (i.e., Individual, Group, Action Types, Module). \( G \) denotes the 6 dimensions in activity theory for each group. \( G_m \) represents the \( m^{th} \) group’s performance in the dimension of Subject, \( G_{2m} \) notates the \( m^{th} \) group’s performance in the dimension of Rules, and so on. For the purpose of this study, all data will focus on the group unit of analysis. In combination with the Data Categorization section, the data set for a single group in this course is a four-fold structure shown in Figure 4:
Measure construction

Subject

Subject in activity theory represents the individual student’s efforts in solving a problem. When mapped to our log data, individual endeavor is a student’s actions over the five event types across all modules where he or she is the starter and receiver. Therefore, for a single student in a group, individual effort can be quantified as:

\[ a^m = \sum_{i=1}^{5} T^m_{ni} \]

Because the Subject dimension in a group is still noted as the individual aspect of endeavor, subjective efforts can first be considered as the accumulation of individual effort from the group members.

\[ A^m = \frac{1}{N} \sum_{m=1}^{N} a^m = \frac{1}{N} \sum_{m=1}^{N} \sum_{i=1}^{5} T^m_{ni} = \]

The average for each group is used here because groups may vary in number of students. Otherwise, groups with more students would have a much higher chance to outperform groups with fewer students.

As a reflection of individual effort in a group, we need to consider whether they perform equally in the modules. Therefore, we use the standard deviation of the group members to show equal contributions:

\[ SD^m = \sqrt{\frac{1}{N} \sum_{m=1}^{N} (a^m - A^m)^2} \]

Therefore, Subject can be denoted as:

\[ G^m_{ni} = \frac{A^m_{ni}}{SD^m} \]

Rules

Rules may be implicit or explicit. Considering that in this course the students conducted themselves through the curriculum, explicit rules have no obvious effect on the group. Under the socio-technological construct, the rules are
the implicit rules that constrain students’ actions, denoted as $T_{act}$. Students can only perform actions that the VMT environment offers.

$T_{act}$ is a set that contains action types in a particular event type. To consider the group performance in Rules aspect, we process group performance over this dimension in each event type separately. First, we put the student members’ *Action Types* over one event type in a unified set. Then for a particular kind of action, for example, “View a Tab” under System, if the students in a group all perform the action, we noted as the number $K_{ij}=N$. Following this logic, if $N-I$ of $N$ students used this “View a Tab” action, it is denoted as $K_{ij}=N-I$. Therefore, for each rule in each event type, the range $K_{ij} \in [0, N]$.

On the other hand, each event type has a different number of rules. To illustrate, Awareness may only contain one rule (Erase the Message) while System has three (Join the Room, Leave the Room and View a Tab). Therefore, combined rule sets over a particular event type may result in a different number. We use $L = \max(|T_{act}|)$ and $n = 1, 2, \ldots, N$ to represent all possible rules contained in a particular event type. Therefore, for any event type $T$, the number of students using one particular rule can be presented as:

$$\{K_{1j}, K_{2j}, \ldots, K_{Lj}\}$$

A simple sum of those numbers could reflect group performance over the rule aspect to some extent. However, it is difficult to differentiate performance among groups because the scale of this summation is too small. For another reason, intuitively, a group of four or five all performing one kind of action requires more effort than a group of three. Thus, we also need to expand the difference among groups with different numbers of group members. We use $e$ as the base to achieve this goal. Therefore, *Rules* is presented as:

$$G_{m2} = \frac{1}{N} \sum_{i=1}^{5} (e^{K_{i1}} + e^{K_{i2}} + \ldots + e^{K_{i5}}) .$$

**Tools**

The *Tools* dimension in activity theory focuses on the process in which tools facilitate group knowledge development. In the VMT context, the tools are System and Wb, where the groups’ actions for tool usage are registered, denoted as $T_{act}$ and $T_{act}$. Hence, for Group $m$, the activity theory *Tools* dimension $G_{m3}$ can be presented as:

$$G_{m3} = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{5} (T_{m1}^{n} + T_{m2}^{n}) .$$

**Community**

Community can be presented as the sum of chat messages and awareness. In the VMT context, students in the same group use chat to directly communicate with each other, and use the awareness function to erase the chat messages which can be categorized as an indirect contribution to the community, denoted as $T_{act}$ and $T_{act}$. Then, for Group $m$, the activity theory *Community* dimension $G_{m4}$ can be denoted as:

$$G_{m4} = \frac{1}{N} \sum_{m=1}^{N} \sum_{i=1}^{2} (T_{m1}^{n} + T_{m2}^{n}) .$$

**Division of labor**

116
**Division of Labor** is a measure of how workload is shared among team members. Labor consists of contributions made by the student to the collaborative learning modules. Although chat messages may also contribute to the development of geometry objects, all concrete contributions to geometry object construction originate from the Geogebra dimension to build the actual geometry objects. Therefore, labor of each student can be denoted as:

$$a_{m5}^i = \sum_{i=1}^{5} \left( T_{a1}^m + T_{a2}^m \right).$$

This dimension would have the highest value if all the members in a group shared the workload equally and would have the lowest value if only one member did the work. Thus, the balance of the work among team members is based on the standard deviation of the group effort with the perfect division $1/N$. However, the smaller the standard deviation value is, the more balanced the effort is among team members. This trend is contrary to the other 5 dimensions. Therefore, we use a minus value to change the trend. In sum, the **Division of Labor** can be represented as:

$$G_{m5} = 1 - \sqrt{\frac{1}{N} \left[ \frac{\sum_{m=1}^{5} a_{m5}^i}{N} \cdot \left( \frac{1}{N} - 1 \right)^2 + \frac{\sum_{m=1}^{5} a_{m5}^i}{N} \cdot \left( \frac{1}{N} - 1 \right)^2 + \ldots + \frac{\sum_{m=1}^{5} a_{m5}^i}{N} \cdot \left( \frac{1}{N} - 1 \right)^2 \right]} .$$

**Object**

The CSCL activity is to achieve the object of a group of students actively participating in the whole class to solve problems. Hence, the first factor to consider is the number of modules in which the group participates. In order to quantify whether the group is active in those learning modules, we incorporate the total frequency of participation and the number of event types. By doing this, we avoid inflated ratings for groups who participate in all modules but make few actions or contributions.

Similar to the Rules function, $T_{a1}$ is a set containing modules in which students participate in that particular event type. Involved in this module is learning object construction. Therefore, instead of a union set $(\cup)$ of all the modules students participate in within a particular event type, we implement a harsher standard intersection $(\cap)$ to highlight its significance. Therefore, for each event type, unless all members in that group are involved in that single module, that module is not counted as a valid number to compute group performance over the object dimension. Then the sum of modules in which all students are involved in over five event types is counted as one aspect of Object.

On the other hand, the number of modules the group is involved in is the key for the **Object** dimension; the other two are secondary factors. While the frequency for participation is in the 1000s, the number of modules is in the scale of 10s, so we use the $\log$ function to dampen the effect of the frequency measure. Although the event types are in the same scale of modules, we still want to dial down for overall event types that the group used. Thus, we use a fraction to lower the effect of event type on object measurement. As a result, each group on **Object** aspect can be denoted as:

$$G_{o6} = \left[ \frac{1}{N} \sum_{n=1}^{N} \log \left( \sum_{m=1}^{5} \left( T_{a1}^m + T_{a2}^m \right) \right) \right]^{\frac{1}{5}} \left[ \frac{1}{N} \sum_{n=2}^{N} [T_{a1}^1 \cap T_{a2}^1 \cap \ldots \cap T_{a5}^1 \cap \ldots \cap T_{a5}^N] \right]^{\frac{1}{5}} \left[ \frac{1}{N} \sum_{n=1}^{N} [T_{a1}^n / 5] \right].$$

In sum, based on activity theory, we have built a quantified model for group performance in CSCL activities specific to the VMT environment as shown in Figure 5: [Subject, Rules, Tools, Community, Division of Labor, Object].
Figure 5. Quantified activity theory model for a group
Group assessment of participation in VMT
Rational
An activity system is characterized by the internal tensions among its components that eventually drive the system to
change and develop, in this context, toward an outcome of group learning. Therefore, it is difficult to compute one
value as functions of the six dimensions to indicate the learning or performance result of a group of students,
especially considering the complexity of the nature of group learning in a technology-mediated environment. In this
exploratory study, we use K-means cluster analysis to chunk groups with similar learning behaviors. A K-means
cluster brings into consideration each of the six dimensions in the activity system rather than accounting for only one
dimension. Another advantage of using cluster analysis is to enable the teacher to understand the overall
collaborative activity performance of the whole class.
Procedure
Cluster analysis, which addresses the problem of data segmentation, belongs to unsupervised learning methods, as
there is no knowledge of “preferred” clusters (Guo & Zhang, 2014). It is a set of techniques used to classify a data
set into groups that are relatively homogeneous within themselves and heterogeneous between each other on the
basis of a defined set of variables (Xing, Guo, Fitzgerald, & Xu, 2014). A significant step in clustering is to define
the system scale and select the proper cluster elements, which we have defined earlier. By considering the
measurements defined in the subsection of measure construction, it is possible to group students into different
categories with a mathematical method. Therefore, the state definition used for this study is a vector of
measurements for all the groups. The data samples are therefore multidimensional because the vectors of measures
for each group are considered simultaneously (six dimensions developed from activity theory). Hence, the system
states in our study are defined as follows, assuming there are 6 representative measurements in the datasets and N
groups. For instance, the constructed measure can be recorded for N individuals (groups). Then the data A , a
K(6)  K(N) matrix, will have the format as follows:

118


To manage differences in scale among variables within each group, the cluster elements should be properly normalized. This process, which uses Eq. (1), is performed before the cluster analysis to reduce dimensionality of the original data.

\[ x_{nm} = \frac{x_{nm} - \bar{x}_{m}}{s_{n}}, \quad (n = 1, 2, \cdots, N; \quad m = 1, 2, \cdots, M) \]

In cluster analysis, cluster elements are grouped according to their similarities, or more specifically, the distances between them (Xing, Guo, Lowrance, & Kochtanek, 2014). Smaller distances between elements indicate greater similarity and higher likelihood of belonging to the same cluster. For our study, squared Euclidean distance, as shown in Eq. (2), is implemented to calculate distance between clusters.

\[ D_{ij}^2 = \sum_{nm} (x_{nm} - x_{jm})^2, \quad (i, = 1, 2, \cdots, N; \quad m = 1, 2, \cdots, M) \]

Where, \( D_{ij}^2 \) is the squared Euclidean distance between measurements \( p \) and \( q \); \( x_{nm} \) is the \( m^{th} \) element in individual \( i \); and \( x_{jm} \) is the \( m^{th} \) element in individual \( j \).

**Results**

**Activity theory based measurements**

Group performance is represented as 6 dimension sets (see Table 2). Results shown in the table are standardized in order to facilitate visualization and tool development in the later section. Through investigation into those numbers alone, a teacher can provide specific advice to a particular group. For example, if the value of a student in the Community dimension (Group 4) is very low, the teacher could suggest that group communicate more among team members. Similarly, if the group has a high value on Subject (Group 6), then it indicates that group members are independently active.

<table>
<thead>
<tr>
<th>Group</th>
<th>Dimension</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject</td>
<td>-0.6293</td>
<td>-0.6667</td>
<td>-0.2755</td>
<td>-0.5633</td>
<td>1.4729</td>
<td>2.7025</td>
</tr>
<tr>
<td></td>
<td>Rules</td>
<td>-0.4625</td>
<td>-0.5365</td>
<td>0.1475</td>
<td>-1.0875</td>
<td>1.2876</td>
<td>1.4486</td>
</tr>
<tr>
<td></td>
<td>Tools</td>
<td>0.1535</td>
<td>0.5637</td>
<td>0.3643</td>
<td>-0.9616</td>
<td>-0.8774</td>
<td>-0.8774</td>
</tr>
<tr>
<td></td>
<td>Division of labor</td>
<td>1.4359</td>
<td>0.3860</td>
<td>-0.5009</td>
<td>-0.3951</td>
<td>1.3323</td>
<td>-1.0588</td>
</tr>
<tr>
<td></td>
<td>Community</td>
<td>1.3411</td>
<td>-0.3055</td>
<td>-0.3404</td>
<td>-0.3951</td>
<td>-0.3491</td>
<td>-0.3491</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td>-0.3078</td>
<td>0.4673</td>
<td>-0.3780</td>
<td>-1.0344</td>
<td>1.8201</td>
<td>1.7011</td>
</tr>
</tbody>
</table>
Cluster results

Using the K-means algorithm, an arbitrary number of clusters are formed for further analysis of results. Table 3 shows final clustering results.

Table 3. Cluster results

<table>
<thead>
<tr>
<th>Measure</th>
<th>Cluster1 (2/14)</th>
<th>Cluster2 (4/14)</th>
<th>Cluster3 (8/14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>-1.343, 2.702</td>
<td>-0.637, 0.560</td>
<td>-0.185, 0.442</td>
</tr>
<tr>
<td>Rules</td>
<td>-1.300, 2.102</td>
<td>1.065, 1.018</td>
<td>-0.321, 0.666</td>
</tr>
<tr>
<td>Tools</td>
<td>-1.034, 2.160</td>
<td>1.367, 0.731</td>
<td>-0.529, 0.330</td>
</tr>
<tr>
<td>Division of Labor</td>
<td>-2.079, 1.449</td>
<td>-0.317, 1.227</td>
<td>-0.172, 0.787</td>
</tr>
<tr>
<td>Community</td>
<td>-1.130, 2.332</td>
<td>0.614, 2.311</td>
<td>-0.041, 0.640</td>
</tr>
<tr>
<td>Object</td>
<td>-1.536, 1.902</td>
<td>-0.217, 0.560</td>
<td>-0.328, 0.837</td>
</tr>
</tbody>
</table>

Cluster 1: Individually participative and great community but less adept in the environment

This cluster contains 2 groups, in which members of these groups were actively involved in the modules as individuals, because they had the highest value (2.012) in the Subject dimension. Similarly, these two groups also ranked the highest in Division of Labor (1.323) and Community (0.614), reflecting that they shared the workload fairly equally and maintain community structure well. Furthermore, they also received the highest value on Object, indicating that these two groups were very active across all modules. By contrast, these two groups received the lowest score on Rules (-0.845) and Tools (-0.616) respectively. Teachers could infer that students in these groups might heavily rely on one or two functions or tools that VMT provides. Therefore, they can encourage students to further explore the VMT environment.

Cluster 2: Great explorer into the environment but not great group learners

This cluster has 4 groups in it and seems to have almost the exact opposite results as compared to Cluster 1. They have the best performance in Rules and Tools measurements, indicating that students in these groups used a variety of tools or functions VMT offers. However, from individual and group learning perspectives, these groups did not perform well enough. In the Subject dimension, they received the lowest value (-0.637) showing that students invested significant personal effort; in the group dimension, students ranked the lowest both on Community (-0.225) and Division of Labor (-0.317) demonstrating that students in these groups did not make equal contributions to problem solving and did not communicate well with group members. This cluster of groups ranked middle in the Object dimension indicating that these groups were relatively active across all modules.

Cluster 3: Average performance as a whole but not consistently active

Most of the groups (8 of 14) fell into this cluster, which has average performance. Subject (-0.185), Rules (-0.321), Tools (-0.529), Division of Labor (-0.172), and Community (-0.041) all ranked in the middle overall groups. In contrast with these dimensions, Object has the lowest value (-0.328). This showed that students in those groups are probably very active in some of the modules but limited participation in others. This is especially true when considered together with the average performance across all other 5 dimensions. While we expect this Object would rank in the middle as well, it countered our expectation. Therefore, teachers may want to discover which module these groups were not active in and check their curriculum design.
Validation and comparison with qualitative chat analysis

To validate our quantitative methodology, qualitative analysis of chat data from Topic 5 was implemented. Topic 5 is mainly designed for students to explore the dependences of geometric objects and build an equilateral triangle inscribed in another equilateral triangle. Using the method described above, these groups have first been put into three different clusters solely based on their collaboration on Topic 5. Then three groups were chosen, one per cluster, and compared with the qualitative analysis results of the three groups. Qualitative chat analysis from an ethnomethodological perspective has been established for years in the context of VMT (Stahl, 2006, 2012), which is a sociological approach to describe the work that people do in collaboration with others to build social order and meaning. In contrast with conversation analysis with a focus on construction of “adjacency pairs,” short sequences in which one person’s utterance ignites a response (e.g., an instance of question-answer pair), ethomethodological chat analysis look for longer sequences, in our context a series of math proposal adjacency pairs (Stahl, 2006, pp. 442-456), of examples of mathematical problem solving by groups (Stahl, 2012). For our chat analysis, we used the VMT Replayer which allows us to see chat and recorded actions at the same time (Figure 2).

Group 14 contains four members (s50, s51, s52, s53) and conversation began with a proposal from a member “Okay lets start by taking turns dragging vertex A and vertex D.” Students in this group took turns proposing actions and building geometric objects, and in the end successfully complete the module objective and are therefore considered a high performing group. This result aligns well with our activity-theory based assessment, which assigns group 6 to the high cluster and the values of 6 dimensions of activity theory are all above class averages. The group as a whole collaborated very well when helping one another use the circle tool correctly in order to make sure the segment they created was the radius of both circles (Figure 6). Specifically, one group member (s50) invested great effort in explaining the process step-by-step for group member (s51) to follow, but she did not succeed. Then s50 helps her teammate, making the circles and intersecting points.

Group 4 has three members (s13, s14, s15) and began by establishing a common ground, agreeing to take turns exploring dependencies in the environment. This group is in the “medium” performing cluster, and although two of the three members came close to completing the assignment correctly, their answer missed key dimensions. Coordination between two of the members went smoothly on moving the vertex of the triangle, but group member (s13) was reluctant in relinquishing control to other members with “Fine” when it is time for others to try. They used correct terms like “equidistant” and “construction” to communicate. Then one of the members proposed to start construction of the two triangles. At the beginning group member (s13) made all the constructions of the triangle without explaining to the other members though other members (s14) asked him to do so “kk tell us what ur doing”. Then group member (s15) asked for control and followed an example taught in class using circles (Figure 7).

Group 12 consisted of four members (s40, s41, s42, s43) and started by moving the vertex, discussing the dependencies they noticed for each triangle. They used many geometric terms but did not typically “notice and wonder,” or follow through on their ideas. This group is from the low performing cluster, and both failed to successfully complete the task, and exhibited many difficulties in collaboration. For example, this group did not...
discuss why they needed to use the Circle function to make the equilateral triangle. One team member (s40) commented “so I think triangle igh is like triangle abc” (Figure 8), again, the team did not discuss why this is so and no member responded to this comment. After they completed the outer triangle, they attempted to build the inner triangle. They talked about how f and e are dependent on d (Figure 8).

In sum, our quantitative analysis obtained similar conclusions with the qualitative chat analysis. Since the proposed method can be totally automated, compared with qualitative method, it is much more efficient and practical in supporting teachers’ decision making because our approach is able to support real time assessment and enable teachers to provide immediate feedback.

**Tool development**

It would be difficult in practice for teachers to infer meaning or make sense of figure 12 without a technical or analytical background. Instructors can more easily obtain information on student participation from these results using visualizations. This is especially true when the number of students is large. The research community in learning analytics has recognized the significance of visualization, establishing it as a required component in the learning analytics cycle (Shum & Ferguson, 2012; Xing, Wu, & Ma, 2014; Xing, Guo, Petakovic, & Goggins, in
press). In addition, a graphical display allows teachers to process information more effectively than do pure numbers (Leinhardt et al., 1990; Goggins, Xing, Chen, Chen, & Wadholm, 2015).

On the other hand, teachers cannot identify specific problems simply based on the clusters of groups. For example, in Cluster 2, teachers may conclude that some students are not as active as others. How can the teacher specifically identify who these less active students are? In Cluster 3, teachers may infer that students are participating in some of the modules but not others, but does not know which modules the students are involved in. A visualization tool would be helpful in this situation. In combination with our previous work on individual assessment (Xing, Wadholm, & Goggins, 2014a), we used graphs to present the quantified 6 dimensional data for both individuals and groups. This allows teachers to see a particular cluster of groups across any module.

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Activity Theory</th>
<th>Cluster</th>
<th>Module</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td>6-dimensional data</td>
<td>3 clusters</td>
<td>Any module or combination of modules</td>
<td>Any group or combination of groups</td>
</tr>
<tr>
<td>Group</td>
<td>6-dimensional data</td>
<td>3 clusters</td>
<td>Any module or combination of modules</td>
<td>Any group or combination of groups</td>
</tr>
</tbody>
</table>

Figure 9 shows the general interface and basic functions of an early draft of the web application, which allows teachers to access this detailed information from anywhere and at any time. With the group, individual, and module information together, the teacher can identify issues with individuals, groups, and the class as a whole in order to provide concrete advice and interventions. Figure 9 (a) shows information on participation across all groups and individuals in 5 modules. It gives an overview of which clusters these groups belong to (High, Average, or Low), the members of these groups, and the clusters each student is individually affiliated with. The six bars with nodes at the top demonstrate information on details of the six dimensions of activity theory for a group. The teacher is able to see how the group performs in each dimension by hovering over nodes in the graph. Since the naming of the six dimensions of activity theory (Subject, Tools, Object, Rules, Community, Division of labor) is relatively abstract, we replaced the six dimensions with Individual Participation, White Board & System, Participation across Modules, Group Participation, Chat & Awareness, and Balance of Workload, for teachers to better understand the results. Hovering over a node shows more detailed information.
Also, there is a button for the teacher to switch visualizations in order to compare the performance of each dimension with the mean for the course as demonstrated in Figure 9 (b). After seeing the big picture of how each group performs, teachers could obtain more detailed information on how each individual in the group performs in Figure 9.
(c). Individual overall performance is identified as being a part of the High, Average, or Low cluster, and further information is shown by clicking on the name of the student. In Figure 9 (d), teachers can click the Switch button to see the radial graph and compare the student’s performance with the mean for the course. On the other hand, teachers could also decide to view the performance of groups and individuals over a particular module, or between specific groups as in Figure 9 (e).

**Discussion & conclusion**

In this paper we explore a different way of developing assessment that is scalable and coherent with theory. Activity theory holistically frames group participation in our framework, which addresses situatedness, contextuality, and social mediation. In addition, activity theory looks at ongoing processes to explain group changes and performance over time (De Laat, 2006). This is important especially considering that a major portion of group assessment in CSCL is still summative in nature (e.g., tests, product evaluation). On the other hand, researchers have created a variety of constructs to assess collaborative processes to infer group learning. These measurements are usually derived from analysis of chat and/or tool usage and therefore unable to assess group learning fully. From a practical perspective, there is an overreliance on text-based measures to assess learning in CSCL (Gress et al., 2010). Coding of the discussions and content are usually quite time-consuming. It is extremely difficult for a teacher using this method to provide timely feedback to individuals and groups. In addition, these coding methods and frameworks are not always shared (Gress et al., 2010) which leads to difficulty in maintaining consistency across different evaluators. The proposed method, from measure construction to clustering and visualization, is totally automated. Therefore, it could significantly reduce the teachers’ assessment burden and offer timely feedback based on assessment results. Automatic methods could also increase the consistency of evaluations and improve the reliability of results.

Though our method is also computational in nature, it is informed by activity theory, which centers on descriptions of human socio-technical behavior. Unlike previous quantitative methods that focus on one aspect of participation such as number of posts, login times or clicks on a page (Wolff, Zdrahal, Nikolvo, & Pantucek, 2013), this theory provides a holistic view of a student’s participation in CSCL activities. While a purely computational method makes it difficult for a teacher to offer tangible advice to students to improve their performance (Xing, Wadholm, & Goggins, 2014b; Xing & Goggins, 2015), activity theory provides a semantic background for the teacher to give specific and meaningful feedback to students. Moreover, many times, computational results are difficult for teachers to use, comprehend and explain (Ferguson, 2012). To facilitate teachers’ timely interpretations of results, we developed a web-based tool that visualizes the data. We are exploring natural language processing of the chat log data and incorporating it into our activity theory measure construction system and visualizations in order to further inform assessment in CSCL and learning analytics (Chen, Chen & Xing, 2015) as a whole.

**References**


Does Digital Scholarship through Online Lectures Affect Student Learning?

Shelley Kinash, Diana Knight and Matthew McLean

Bond University, Office of Learning and Teaching, Bond University QLD 4229 Australia // Bond University, Office of Learning and Teaching, Bond University QLD 4229 Australia // Australian Government, Department of Social Services, Australian Government ACT 2610 Australia // skinash@bond.edu.au // dknight@bond.edu.au // matthew.mclean@live.com.au

*Corresponding author

(Submitted August 15, 2013; Revised November 29, 2013; Accepted August 19, 2014)

ABSTRACT

University lectures are increasingly recorded or reproduced and made available to students online. This paper aggregates and critically reviews the associated literature, thematically organised in response to four questions. In response to the first question - does student attendance decrease when online content is made available - research indicates that students primarily use digital content for review and revision rather than as a substitute for on-campus attendance. Analysis of the research in response to the second question - is achievement affected when attendance is face-to-face versus online - revealed no empirically supported significant difference. The third question was whether online content is better suited to some pedagogical tasks than others. A predominant theme in the literature is that digital content has potential as a disruptive pedagogy, accelerating an overall shift from didactic lecture to constructivist learning. Analysis revealed a research gap around the fourth question - is there evidence that some online formats are particularly suited to advancing learning. The few published comparative studies revealed contradictory results. Overall conclusions from the combined questions are that online digital content is a worthwhile learning and teaching pursuit and discipline and context must be considered in designing the particular approach.

Keywords
Digital scholarship, E-learning, Lecture capture, Lecture recording, Online digital content

Introduction

Digital scholarship refers to connected and blurred knowledge producers and consumers, creating, sharing and constructing learning with a global network of peers (Losoff & Pence, 2009; Markauskaite, 2010; Weller, 2012). Within universities, a critical (and critiqued) aspect of digital scholarship is providing recorded lectures for online delivery (Newbury, Watten, Holroyd, & Hardman, 2011). Some lectures are recorded while they are being delivered in a theatre or classroom (lecture capture) and then posted online (Al Nashash & Gunn, 2013). Lectures are also produced for online dissemination from computers outside of teaching spaces and rendered in various formats such as podcasts and audio-narrated slideshows, often referred to as web-based lecture technologies (Germany, 2012). Whether and how lectures should be provided online has been a heated debate in higher education (Cardall, Krupat, & Curro, 2012; Gulseth, 2008; von Konisky, Ivins, & Gribble, 2009). There are three primary factors for consideration in support of providing students with online lectures. First, students are requesting online delivery (Copley, 2007; Evans, 2008; Jensen, 2011). Second, education technology offers opportunities to make online content viable and many educational theorists argue that web-based technologies have learning advantages (Grabe & Christopherson, 2007; Holbrook & Dupont, 2009; Lonn & Teasley, 2009; McGarr, 2009; Nworie & Haughton, 2008). Third, educational institutions struggle to continue providing quality learning opportunities, while managing the rising costs of providing them, in the context of knowledge management (Brewer & Brewer, 2010; Guzman & Trivelato, 2011; Omona, van der Weide & Lubega, 2010; Sulisworo, 2012; Yuan & Powell, 2013).

Universities and academics are paying heightened attention to digital scholarship in the form of online lectures, explained in large part by the entry of massive open online courses (MOOCs) as a disruptive pedagogy (Cusumano, 2013; Hyman, 2012; Pence, 2013). As costs of higher education rise and MOOCs increasingly occupy a larger share of on-demand education, there is a sense of urgency and trepidation that educators must adapt and implement digital content and recorded lectures, without fully comprehending the significance (Van Zanten, Somogyi, & Curro, 2012; Yuan & Powell, 2013). Interwoven with pedagogical and value-based decision making, emerging innovations in educational technology including third-party provider products and services are streamlining processes to create and/or record content, thereby fostering efficient, effective and often interactive online learning environments (Wells...
& Brook, 2008; Yuan & Powell, 2013). Although there is a long and rich history of distance education research (e.g., Moore, 2013), the emphasis on MOOCs has renewed attention on student access to learning materials and to study beyond the classroom. Educators are investigating the pedagogy of MOOCs and asking questions about the changing roles of teacher and learner, digital engagement, and balancing between openness and control (Bayne & Ross, 2014). Although MOOC pedagogy research is still in its relative infancy, a critical review of lecture capture may help shape the foundations upon which extended MOOC and distance research may stand. Therefore, while lecture capture has a relatively long history, it is an emergent priority for educators and education leaders.

Universities use lecture capture and streaming and some are beginning to use web-based lecture technologies as a primary means of creating and disseminating digital scholarship. The concerns that are emerging from academics and the content of documented investigations cluster in response to four questions:

- Does student attendance decrease when online content is made available?
- Does it matter if student attendance decreases?
- In other words, does achievement decrease if, and when, student attendance decreases?
- Is online content better suited to some pedagogical tasks than others?
- Do some types of online content work better than others?

**Methodology**

There are a large number of publications addressing topics and issues about lecture capture, web-based lecture technologies and online delivery as a form of digital scholarship. Currently, informing decisions about whether and how to capture or produce and distribute lectures requires sifting through numerous abstracts and papers. There appear to be no systematic reviews of the literature and analyses of the overall content, outcomes and conclusions. In response, this paper presents collation of numerous papers and book chapters about online lectures, attendance and student achievement. The research gap addressed by this paper is one of collation, synthesis and application. While the topic of online lectures is reasonably well represented in the literature, the studies are set at individual institutions and within specific disciplines. The gap is a straight-forward response to higher education leaders and academics about whether or not to incorporate online lectures into their pedagogical approach and if so, how to proceed. Gunn and Steel (2012) wrote, “strong, tested and connected evidence that theory informed, technology-enhanced designs can improve learning outcomes is required to further reduce the historical gap between educational research and practice” (p. 11). Theory develops through evaluating, disseminating and exploring synergies, commonalities and relationships between individual, singular, snap-shot studies. The contribution that this particular paper makes to the literature is drawing together individual studies to look for commonalities and inform theory.

The literature was classified to inform a response to each of the four questions. Responses were then combined to provide recommendations as to whether (and how) to proceed with provision of digital scholarship. The methodology of this paper is research synthesis as a process of collecting, collating and analysing the published literature to advise next steps in application and further research (Cooper & Hedges, 2009). The scope of this review includes content on lecture capture, web-based lecture technologies and distribution. Other means of creating and disseminating digital content (e.g., wikis, blogs) have been omitted.

**Table 1. Keyword search results**

<table>
<thead>
<tr>
<th>Term used (2005-2013)</th>
<th>Search returns (no.) Google Scholar</th>
<th>Search returns (no.) Ebsco Megafie Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>lecture recording (LR)</td>
<td>75,400</td>
<td>228</td>
</tr>
<tr>
<td>university LR</td>
<td>68,700</td>
<td>3</td>
</tr>
<tr>
<td>university LR, attendance</td>
<td>16,000</td>
<td>0</td>
</tr>
<tr>
<td>university LR, achievement</td>
<td>18,100</td>
<td>0</td>
</tr>
</tbody>
</table>

The initial strategy to identify journal papers was a search of Google Scholar and the electronic database, Ebsco Megafile Complete. An initial sampling without specifying the publication year revealed that the content of the papers prior to 2005 were not relevant because of the types of educational technology employed. As indicated in Table 1, there were substantially more records in Google Scholar, but the majority either did not meet the criteria of peer-reviewed empirical journal papers and book chapters, or were duplicate entries. The electronic database
revealed substantially fewer results. There were zero electronic database returns when combining the search terms, *university, lecture, recording* with either *attendance* or *achievement*. The database uses SmartText searching to source additional results based on the submitted keywords. Most of these papers revealed small mentions as a sub-theme in a paper on a different research theme. A manual search of key educational technology journals and bridging from the relevant papers’ end-text references revealed additional sources. Three coders identified the key scholarly papers with thematic relevance resulting in a detailed analysis of 30 publications.

**Combined themes**

Of the 30 publications that addressed one or more of the questions regarding online content, 19 of the papers described a research project whereby some evidence was collected and analysed to address the effect of providing online content, either on student attendance, student achievement or both of these factors. The other 11 are addressed under the heading of each of the paper’s four main questions.

<table>
<thead>
<tr>
<th>First Author &amp; Year of Publication</th>
<th>Empirical (Experimental)</th>
<th>Lecture Capture</th>
<th>Attendance</th>
<th>Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Al Nashash, 2013</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Billings-Gagliardi, 2007</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Bongey, 2006</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Cardall, 2008</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Copley, 2007</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>DiVall, 2013</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Evans, 2008</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Grabe, 2007</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Holbrook, 2009</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Jensen, 2011</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Lents, 2009</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Lewis, 2012</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Lonn, 2009</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>McKinney, 2009</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Nast, 2009</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Traphagan, 2010</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>von Konsky, 2009</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Wang, 2010</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Williams, 2012</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 2 presents an analysis of the reviewed literature. The first column identifies the papers by the surname of the first author and year of publication. The full citations are available in the end-text references. The second column classifies the methodological design of the papers into quasi or fully experimental or other - usually conducted through survey. The third column depicts whether or not the paper addressed digital scholarship through lecture capture or some other format. The fourth column identifies whether the research results showed a positive, negative or neutral effect of digital content on student attendance. A small subset of papers did not address attendance, in which case, there is no x inserted in this set of rows. Similarly, the final column identifies whether the results showed a positive, negative or neutral effect on student achievement. Again, where no x is inserted, achievement was not addressed.

Nineteen journal papers were found that researched the impact of digital scholarship on the higher education student experience. These papers were published between 2006 and 2013 with the largest number (six papers) appearing in 2009. The methodological design of ten of the papers was experimental or quasi-experimental and the other nine were survey-based. Ten of the papers researched online content created through lecture recording. The other nine addressed other types of online formats, the most prevalent of which was podcasts. Eleven of the papers found no evidence in support of the hypothesis that providing lectures online increases student absenteeism. Results of three of the studies showed that there was a deleterious impact. The remaining five did not address attendance. Eleven of the
papers provided evidence that online lectures positively affect student learning or achievement, while five studies showed no effect. Three papers did not address student learning and three did not address achievement. Analysis of the literature led to the following overall conclusions. Approximately half are experimental in design and approximately half address lecture capture. The weight of the evidence is that providing lectures online does not decrease student on-campus attendance and that it increases achievement.

The approach and results of the survey studies are now briefly presented to provide further information regarding perceptions about the relationship between digital scholarship, attendance and achievement. In survey findings, Billings-Gagliardi and Mazor (2007) and Bongey, Cizadlo and Kalnbach (2006) revealed that the availability of digital materials did not decrease their on-campus attendance. Cardall, Krupat and Ulrich’s (2008) student survey revealed that while the majority of learners continue to attend live lectures when provided both options, those who access recorded lectures do so because they believe their learning is improved. Students in Copley (2007) reported using the online materials for revision, and not decreasing their on-campus attendance. Copley interpreted the survey data as providing evidence of enhanced learning outcomes. He explained that by supplementing students’ online study resources, they were freed-up to engage and participate rather than to record notes in class, thereby scaffolding thinking and activity (Bransford, Brown, & Cocking, 1999). Research findings also indicated a high percentage of surveyed students indicated a belief that lecture capture and podcasts improved learning (DiVall et al., 2013; Evans, 2008, even when students reported decreased attendance (Holbrook & Dupont, 2009). Faculty members were also positive about lecture capture for improved learning, but to a lesser extent (Lonn & Teasley, 2009). The students surveyed by Wang, Mattick and Dunne (2010) perceived face-to-face lectures as more consistently maintaining quality standards and stated that absenteeism would not increase as a result of online content. The reviewed survey literature revealed that the research participants were most often students and that the most common response was favourable to digital scholarship.

Next, the results of the experimental and quasi-experimental research are described to annotate the evidence regarding an overall positive relationship between digital scholarship and achievement and detail how researchers established no significant relationship with student on-campus lecture attendance. Grabe and Christopherson (2007) found that when given options of type of online content, most students used complete text-based lecture notes. Absenteeism did not increase. Notably, there was a positive relationship between student attendance and use of digital resources. Further, this positive relationship extended to achievement as demonstrated through exam performance. Jensen (2011) structured an experiment varying the pedagogical approach in alternate weeks. The attendance at on-campus lectures and viewing of online lectures decreased and quiz scores indicated no significant difference for either approach. The students in this cohort, however, perceived face-to-face lectures as being more advantageous for their learning. Lewis and Sloan (2012) found a slight, but statistically insignificant benefit of online lecture viewing on quiz performance. The researchers continue to explore means of increasing the impact. Williams, Birch and Hancock (2012) provided empirical evidence that some students use captured lectures as a replacement for on-campus attendance and others as a supplement or revision tool, thereby attendance is indicated in Table 2 as neutral. Students who used online lectures as a complement to face-to-face classes achieved higher grades. While a full meta-analysis is not possible due to limited commonality and standardisation between methodologies and factors, common conclusions as to unsupported links between digital content and absenteeism and supported relationships with learning were established.

Four of the analysed papers described research designs such that online lectures were treated as the independent, experimental, causal variable and attendance and/or achievement were treated as the dependent variables, or effect. In the context of introductory biology, Lents and Cifuentes (2009) compared two sections of the same course, whereby 59 students experienced all of their lectures face-to-face and 24 students experienced 8 of their lectures online through screen-casting (adding audio narration to slides and producing as an online video). The outcome was no significant difference in student achievement or attendance in on-campus classes. Studying a single undergraduate software engineering subject with 108 students, von Konsky, Ivins and Gribble (2009) tracked attendance, grades and student access to streamed lectures. Consistent with Lents and Cifuentes (2009), there was no indicated relationship between online lectures, achievement and attendance.

In McKinney, Dyck, and Luber (2009), 32 undergraduate psychology students experienced on-campus lectures and 34 students experienced podcasts with accompanying printed slides. The podcast group showed statistically higher exam scores. Notably, the students self-selected research groups and even though the researchers analysed and found no significant difference in overall grade point average between students in the two groups, the sample size was too
small to overcome potentially interfering variables. Based on class section registration, Traphagan, Kucsera and Kishi (2010) conducted research with 211 geology students who attended face-to-face lectures and 153 on-campus students who also had access to the lectures online, wherein the slides and lecturer played on side-by-side windows. The results indicated reduced lecture attendance by the group with access to online recordings. However, there was no significant difference in achievement between the two groups and there was a positive relationship between online lecture viewing and achievement. As with other presented studies, these researchers did not randomly assign students to research groups.

In summary, the results of the four published empirical studies reviewed above are unanimous in revealing that student achievement is not impaired by having access to online lectures. The studies warrant further investigation into a hypothesised positive relationship between digital scholarship through online lectures and student achievement.

**Attendance**

The primary articulated question in the online lecture debate is whether attendance decreases when digital content is made available to students. The position that lectures should not be posted online is primarily advanced by the concern that if digital content is made available, students will no longer come to class (Billings-Gagliardi & Mazor, 2007; Romanelli, Cain & Smith, 2011; van Zanten, Somogyi, & Curro, 2012). Analysis of publications on the topic of online lectures and student attendance revealed that survey methodology was the primary empirical approach employed. In short, students were asked whether they would come to class and whether they do attend when lectures are provided online. The compiled strong majority response from students is that they will and do continue coming to class even when lectures are also provided online. Despite large sample sizes, due caution is warranted in that student self-report often results in “recall and social desirability biases” (Cardall, Krupat & Ulrich, 2008, p. 1178). In other words, whereas students reported that they came to class, their attendance was largely not confirmed.

Even with methodological caution, the evidence is compelling in that there is significant agreement between multiple studies that students who have access to online lectures will continue to come to class (Al Nashash & Gunn, 2013; Billings-Gagliardi & Mazor, 2007; Bongey, Cizadlo & Kahlbach, 2006; Brittain, Glowacki, Van Ittersum & Johnson, 2006; Nast, Schafer-Hesterberg, Zielke, Sterry & Rzany, 2009; Wang, Mattick & Dunne, 2010). Each of these studies queried the relationship between student reported attendance and access to recordings or reproduction of lectures. All of the studies reported an insignificant relationship between the two variables. In other words, the argument that lectures should not be provided online because students will stop coming on-campus is largely unsubstantiated.

**Achievement**

The focus of the debate regarding whether to provide lectures online in digital formats may be misplaced. The follow-up question is whether student attendance at lectures matters. If student attendance were lower in classes where the lectures are provided online, would student achievement (i.e., grades) also be lower? Is there a statistically significant relationship between in-class attendance and student learning outcomes? The focus on learner presence in face-to-face classrooms makes it “easy to forget that student achievement in school also depends on what happens outside of school” (Bransford, Brown, & Cocking, 1999, p. 224). Multiple authors in the analysed literature wrote that the decision about lecture recording should be made on the basis of student learning rather than student attendance (Lents & Cifuentes, 2009; Romanelli, Cain & Smith, 2011). E-learning theory supports the concept that neither learner social presence nor cognitive presence is sacrificed when lectures are online (Garrison & Anderson, 2003). The number of students present in class is not a valid performance indicator if attendance is not a statistically significant proxy for learning. Therefore, before addressing the question of whether attendance in live lectures versus online lectures matters, it is important to consider whether it is assumption or proven that any type of attendance (defined as scholarly presence, face-to-face or online) affects student achievement.

Some published studies have probed the relationship between attendance and achievement. Dollinger, Matyja, and Huber (2008) posed the question through quasi-experimental research with 338 undergraduate students. Their results
indicated that attendance combined with study versus work hours accounted for only 6-10% of the variance in exam scores. In other words, even when the factors of attendance and out-of-class study time were combined, these factors made a small difference to student grades. Conversely, Fox and Medhekar (2010) reported a strong relationship between attendance and achievement. A regression model run on an undergraduate macroeconomics subject predicted a “25% increase in performance for students with high attendance compared with those with low attendance” (p. 98). Jensen (2011) found no significant difference in achievement between two groups of 115 introductory psychology students experiencing the lectures face-to-face or through video lecture on a rotating schedule. Although some studies have produced empirical support for the argument that students who attend lectures achieve higher grades, this is contradicted by a number of other studies. The results may indicate that some studies have effectively harnessed the potential of technology to enhance learning out of class, whereas other studies used the tools as an add-on. Technology, of course, does not guarantee effective learning and the relationship between student attendance and achievement is far from being confirmed and warrants further research.

**Asking students whether online lectures affect learning**

The literature reviewed above failed to establish a strong link between any type of lecture attendance (online or face-to-face) and student achievement. Other studies, conducted through surveys, queried whether students believe there is a relationship between online lectures and achievement. A UK survey revealed that 74% of student respondents were of the belief that information technology was very useful in enhancing their learning (Ipsos MORI, 2008). Similarly, an Australian survey found that lecture streaming was perceived to help 68% of students in a significant or moderate way to achieve better results and 80% said that online lectures made it easier to learn (Flores & Savage, 2007). Moreover, the results of multi-university surveys and case studies conducted by Gosper et al. (2008) described a number of learning advantages to online lectures including: use of web-based learning technology for exam revision, clarification of complex concepts, control of one’s study pace and place, and opportunities for comprehensive notes and review prior to approaching the lecturers. In summary, the literature did not support an empirical relationship between in-class attendance and student achievement. Surveyed students, however, believe that a positive relationship does exist between online lectures and their achievement and/or their learning process.

**Pedagogy**

The third question emerging through secondary analysis of the literature is whether online content is better suited to some pedagogical tasks than others. A common theme in the analysed literature is that face-to-face and online pedagogy serve different learning and teaching purposes. Authors argued that lectures can be effectively distributed online, whereas labs, tutorials and classroom activities are better facilitated on-campus (Brittain, Glowacki, Van Ittersum, & Johnson, 2006; Lents & Cifuentes, 2009; von Konsky, Ivins, & Gribble, 2009). Gump (2006) articulated the obvious pedagogical observation that a participative approach to learning and teaching is only possible if students are present. However, there is heightened awareness, complimented by developments in technology-enabled learning, that presence is not restricted to face-to-face (Garrison & Anderson, 2003; Wei, Chen, & Kishuk, 2012). Sound learning designs through prodigious use of information and communication technology in teaching will support quality learning outcomes (Lam, Chung, & Lam, 2010; Nisar, Munir, & Shad, 2011; Oliver, Harper, Wills, Agostinho, & Hedberg, 2007). For example, Odhabi and Nicks-McCaleb (2011) reported the use of video cameras and microphones set to record and distribute classroom activities from both the professor’s and students’ visual points of view. The authors presented evidence that this application of technology-enhanced learning appeared to improve student understanding. A predominant theme in the literature was that a blend of online and on-campus pedagogical tasks contributed to a well-rounded student-centred experience (Demetriadis & Pombortsis, 2007; Odhabi & Nicks-McCaleb, 2011; Romanelli, Cain, & Smith, 2011). Authors addressing the blended learning theme discussed the potential for digital content to be transformative, such that reflecting on, designing, creating and distributing and evaluating digital scholarship has the potential to change teaching conceptualisations and approaches from didactic lectures to constructivist learning. Authors emphasised that learning should be the constant guide of what and when technology can serve as the vehicle through which teaching is facilitated.
Format

The fourth question is whether some types of online content work better than others. In particular, this question asks which education technologies (e.g., videos, podcasts) are more effective in improving student learning. Some of the analysed studies experimentally compared multiple approaches. Brittain, Glowacki, Van Ittersum, and Johnson (2006) analysed 70 survey responses from dentistry students. One of the questions asked students to indicate their preference between podcasts that were synced with PowerPoint slides and video podcasts (vodcasts) made by recording lectures. Of those who indicated using the media, the majority (66%) preferred audio-only. Follow-up inquiry indicated that the primary reason given for the preference was mobility. Similarly, Copley (2007) studied comparative uptake and survey-derived student preference (n = 84) for audio-only podcasts versus slides with audio narration saved as video files. More students downloaded the audio-only files (80%) than the video files (61%). However, student indicated preference was slightly higher for the video files (4.7/5) than for the audio-only files (4.4/5), with 91% of students indicating that video format “provides a complete record of the lecture” (p. 393). Grabe and Christopherson (2007) conducted research with 329 introductory psychology students who had access to online content through: (a) slides only, (b) lecture notes and (c) lecture recordings. The download data revealed that 61% of students accessed the slides, 19% the lecture notes and only 3% the lecture recordings. The researchers theorised that students perceive written notes, and particularly those of a compressed or summary nature, as more efficient than listening to an entire lecture. In a more recent study, Jadin, Gruber, and Batinic (2009) asked whether the design of online lecture content affected learning, or whether student learning styles were the principal determinant in achieving learning outcomes. The researchers provided both multi-modal e-lectures containing text, video and links to additional resources and unimodal lectures that did not contain synchronised text. The authors concluded that the additional text of the oral presentation did not affect learning performance. Instead, students’ measured preferences for certain learning strategies were a stronger factor affecting learning. The authors identified two main types of learners; “repeaters”, who watched the same lectures multiple times and “surfers,” who tended to access additional links for supplementary materials. “Repeaters” outperformed “surfers” when tested (Jadin, Gruber & Batinic, 2009).

Discussion and conclusions

The research established that student attendance does not seem to decrease when online lectures are provided, and that it does not appear to affect student achievement whether they observe lectures live or online. Many authors concluded that face-to-face and online formats are only equivalent when used for didactic information that can be delivered as a lecture. Students require opportunities for experimentation and intellectually rich environments for discussion, debate and Socratic questioning. Among the analysed literature, there was strong support for the proposition that these pedagogical tasks are best facilitated face-to-face. Meanwhile, education researchers also described digital scholarship as a disruptive innovation, in that it can lead to imagination and renewal in the learning and teaching experience.

The rationale for many of the analysed papers was an empirical inquiry into the lay-logic that if lectures are posted online, students will not come to class. Now that this premise has been empirically unsubstantiated, further research may move on to consider why we would consider posting digital content. As evidenced in the review, digital content can be rendered in alternative formats and is therefore accessible to those students with diverse needs such as sensory impairments. Online files can be reviewed before class as pre-study, after class to confirm understanding, and prior to exams as a study strategy. Online files allow pause, fast-forward and rewind functionality to accommodate each learner’s unique pace. The review indicated that this is particularly advantageous to students with diverse learning needs.
Throughout the analysed literature, there was minimal specific reference to educational theory. This is a common feature of educational technology research (Gunn & Steel, 2012). However, considering the body of literature as a whole, theory begins to emerge. In the context of educational technology, Gunn and Steel (2012) defined theory as “an organizing framework that brings an additional layer of understanding to concrete experience by implying relationship, consistency and a degree of predictability and testability” (p. 8). The literature that addressed online lectures was concrete, practical and practice-oriented. However, from this practical base, features of a framework emerged.

- Learning is the desired outcome and technology is the enabler or enhancer.
- Students are important stakeholders in the higher education context and their perceptions are valued.
- Diversity is acknowledged and it is appreciated that solutions must be developed within particular disciplines and with unique student cohorts.
- Academics are teaching researchers, evaluating, collecting evidence, reflecting and revising their approach.

While these propositions are appealing in their straightforward, linear nature, they are overly simplistic. Education is messy. There are no clear independent and dependent variables, or causes and effects and it is a fallacious deterministic notion to state that technology can enable learning (Selwyn, 2012). Selwyn (2012) challenged readers to move beyond questions such as whether online lectures improve learning to inquire into the broader social contexts and power (im)balances that provoke these issues. It is important to challenge pragmatic interpretations and applications of educational technology theory with critical social theory to examine the consequences of our choices and actions and be clear and transparent about our intentions (Hall, 2011).

**Recommendation**

The overall intention of this literature review was to recommend whether universities should provide lectures online through digital content. Research across disciplines and in numerous universities worldwide revealed that the benefits of online lectures outweigh the disadvantages. Some specific recommendations accompany the advice to proceed. For learners, in addition to the provision of lectures, universities are encouraged to provide accompanying explicit study strategies. The expectations and purpose for digital content must be communicated to students. For educators, the literature recommends intentional pedagogy and development of strategies for creating lectures that are effectively disseminated online. The creation of digital content must have a clear teaching and learning purpose that aligns with the module’s learning objectives. It must not be a mere add-on with no purpose. Digital scholarship can enable learners and teachers to experience presence and higher order learning, accomplished through moving beyond the dissemination of online content to linking, creating, evaluating and developing. The overall theme throughout this review was that the university must consider the profile of the educators and learners in the specific context.

**Future directions in research**

Debate about online lectures provides an opportunity to re-examine why students enrol in university and what teaching approaches best support their learning. Further research is required as to how to design digital content to heighten student interaction in face-to-face and online contexts. There is ample room for further inquiry. For example, one of the questions that does not appear to have been empirically addressed is – in the context of social media, will students hesitate to ask questions and participate in discussions in-class if the lecture and thereby their contributions are being captured? It is anticipated that this review of the literature and the results of the intended study will contribute toward improved understanding and an insight into the design and process of technology use for student achievement. Other remaining unanswered questions include whether and when it is preferable to produce online lectures by recording regularly scheduled on-campus lectures or producing stand-alone segments on a separate occasion either from one’s own personal computer or in a university-based studio. Further, what are the boundaries around intellectual property and privacy concerns? There is a substantive body of literature establishing the efficacy of digital content. In the face of increasing emphases on MOOCs, blended learning strategies, and flexible delivery, the next step is to inquire into the design process and continue empirically querying the relationship between digital scholarship and student learning.
References


A Two-Stage Multi-Agent Based Assessment Approach to Enhance Students’ Learning Motivation through Negotiated Skills Assessment

Abdelhafid Chadli, Fatima Bendella and Erwan Tranvouez

Computer Science Department, Ibn khaledou University of Tiaret, Algeria // Computer Science Department, University of Sciences and Technology of Oran, Oran, Algeria // Data, information & content management group, Aix Marseille University, CNRS, LSIS, UMR 7296, Marseilles, France // chadli.abdelhafith@univ-tiaret.dz // fatima.bendella@univ-uosto.dz // erwan.tranvouez@lsis.org

ABSTRACT
In this paper we present an Agent-based evaluation approach in a context of Multi-agent simulation learning systems. Our evaluation model is based on a two stage assessment approach: (1) a Distributed skill evaluation combining agents and fuzzy sets theory; and (2) a Negotiation based evaluation of students’ performance during a training simulation or a Problem-solving process in a Computer-assisted learning system. This paper highlights how this approach deals with the problem of subjective evaluation of students, and shows the impact of Negotiated skills evaluation on reducing the students’ rate of dropout. This approach can also compensate the absence of human expert for assessing training results. Applied to training in plant protection, experiments' results showed first the fuzzy sets based assessment to be similar to the domain expert’s assessment and second the negotiated skills assessment to be effective in assessing students’ abilities and sustaining students’ motivation to continue learning. This evaluation approach allows us to address the problem of subjective assessment and overcome some difficulties encountered in traditional measurement models.

Keywords
Agent-based evaluation, Negotiated collaborative evaluation, Distributed evaluation, Fuzzy logic based assessment

Introduction

The use of Simulation based systems for education and training purposes is still hindered by lack of methods and tools to assess learners’ progress during a training session. For instance, in classroom-based learning, assessment is usually conducted in two ways (formative and summative) and is performed by human experts. However, in Simulation-based learning, these assessment methods become inappropriate, as they often consist in a negative feedback without explanation or improvement guidance, which can lead a learner to lose motivation and to stop learning. Furthermore, when it comes to assessment, there is no appropriate Computer-based assessment methodology adapted to Simulation-based learning/training (Ekanayake et al., 2011). Currently, skills assessment in training simulations is often conducted by human instructors using subjective qualitative methods (based on human expertise), which becomes difficult to automate as expected in Simulation-based learning systems in regards to reduction of instructional time and costs (Eck, 2006).

In order to help students better cope with difficulties encountered in solving problems, many researchers have developed intelligent assessment tools based on artificial intelligence approaches (Stathacopoulos, 2005; Huang, 2008). For example, the conceptual framework developed by Mislevy et al. (2003) adopts an Evidence-Centered Design (ECD), which informs the design of valid assessments and can yield real-time estimates of students’ competency levels across a range of knowledge and skills. However, the following issues in existing assessment models require further investigation:

- Assessment tools often proceed in a single stage evaluation of student’s skills, and focus more on producing marks than giving detailed explanation on what the students failed to understand or put into practice (Chang et al., 2006). Furthermore, learner feedbacks may be insufficient and lack accuracy to help students.
- Generally, all assessment tools set a threshold score for tests to be succeeded. This discriminates students whose final score is near the passing limit. Does a student with a 9.9 final score have significantly less knowledge than a student with 10 as final mark? Moreover when considering potential error margins.
- The existing assessment tools focus more on assessing learner’s performance regardless of whether this assessment contributes to learners’ motivation and not give up of learning.

These issues can be addressed first by refining the assessment skills criteria in order to detail what part of the learning process went wrong. Secondly, marks should be handled with a margin error, thus avoiding threshold phenomena where an assessment can change significantly. Moreover, taking into account limited compensation...
between the different assessment criteria (as when deciding whether a student should graduate or not), would grant a more flexible assessment as humans do. Finally, reporting feedback to the student can then be detailed and not fully negative, reducing the demotivation issue.

In this paper, we propose to use distributed assessor agents to assess skills individually and thus inform precisely about difficulties encountered by the student at each skill. By using Fuzzy sets, assessor agents are able to evaluate the level of control of each skill by considering the difficulty of each action of the skill. Our strategy involves a two-stage approach (see figure 1): The first stage focuses on student’s skill evaluation by means of assessor agents; each is responsible of evaluating only one skill of the student. This will inform the second stage of the approach, which concerns the global evaluation of the student’s capabilities. This evaluation stage is managed by an aggregate agent and is based on the assessor agents’ assessments, allowing a negotiation process to decide whether the student passes the required skills qualification.

The proposed system was evaluated by conducting three experiments using students training in plant protection as subjects. The following three issues are explored here:

- Comparing the students’ assessment by the system to an expert like.
- Whether provided feedback helped students with their problem solving.
- Whether assessment method encouraged students to continue learning and not give up the learning system sessions.

This paper is structured in four parts. We first present relevant research works in learning assessment area, related issues and the problematic we address. Then we describe our skills evaluation approach, in this part we detail firstly the skill evaluation and secondly the global evaluation where we expose all related aspects to the negotiation process. Finally, we present the experimental design and analytical results before drawing the conclusion.

**Relevant research works and related issues**

A number of intelligent learning environments are concerned with evaluating learners in a purpose of providing individual learning guidance to learners, and enhancing the performance of each learner. For example, Stathacopoulou et al. (2005) have developed a neural network implementation for a Fuzzy logic-based model of the diagnostic process as a means to achieve accurate student diagnosis in intelligent learning environments. Huang et al. (2008) have proposed an intelligent diagnosis and assessment tool, and have incorporated it into an open software e-learning platform developed for programming language courses, based on text mining and Machine learning techniques to alleviate teachers’ workload. In order to address the question: How can Game-based learning be assessed? Shute and Ke (2012) have adopted a form of stealth assessment based on ECD-based assessments. All these works led to powerful tools that circumvent the weaknesses of the traditional assessment. However, these assessment tools did not deal with the issues referred to above.

Distributed skill centered assessment is a method that can address these problems by combining heterogeneous skill assessment methods and thus enable the development of robust and valid simulation or Problem solving based learning systems (Oulhaci et al., 2013). Additionally, the presence of imperfect information is an important factor which often leads to errors in the learner evaluation. As Fuzzy sets theory is classically used to handle uncertainty by using qualitative output variable, using the learner’s level over a knowledge classification scale can help solving this limitation. Many studies focusing on assessment have used the Fuzzy theory, for example, Innocent et al. (2005) have developed a Fuzzy expert system for the medical diagnosis to show how fuzzy representations are useful for taking uncertainty into account and can be applied to model the acquisition of knowledge by experts. In our assessment tool, the first stage’s assessor agents are in charge of applying the Fuzzy logic process to evaluate each skill master degree.

In this research, we also promote the use of Negotiated collaborative assessment based on Fuzzy sets individual evaluation. Some researches proved that discussion and negotiation between independent assessors can enhance the reliability of assessment criteria for portfolios (Rees & Sheard, 2004). According to Pitts et al. (2002), the collaborative assessment makes it possible to provide an enhanced assessment of learner so as to improve his skills. This idea was used first in open model of the student's knowledge (Dimitrova et al., 1999), which involved the student and the assessor (or an agent acting on the assessor's behalf) negotiating an agreed assessment of the student
Negotiation is also used in peer assessment, Lai and Lan (2006) have developed a method for modeling collaborative learning as multi-issue agent negotiation using fuzzy constraints. The proposed method aggregates student marks to reduce personal bias. In this framework, students define individual Fuzzy membership functions based on their evaluation concepts and agents facilitate student-student negotiations during the assessment process. In the same context, several authors show that the allocation of bonuses to students is a relevant intervention strategy for cognitive engagement and student achievement (Black & Duhon, 2003; Tuckman, 1996).

The approach that we propose allows learners to engage in a Simulation based learning and offers the opportunity for the learner, whose final score is close to the threshold fixed for test success, to negotiate his final score by means of assessor agents through the use of the results of this learner’s skills assessments and bonus attribution according to the student’s skill performance.

Agents have been widely used in training environments for different purposes. Multi agent modeling is probably the most used tool in these environments, notably by using their intelligence ability and autonomy for learner’s assessment. For example, Baghera is an intelligent tutoring system which uses a theorem prover agent for automatic verification of proofs (Caferra et al., 2000). Pilato et al. (2008) have used a conversational agent to assess the student knowledge through a natural language question/answer procedure. They have used the latent semantic analysis technique to determine the correctness of the student sentences in order to establish which concepts the student knows. An assessment agent is used by Lai and Lan (2006), this approach allows student, whose coursework is marked, negotiate with markers using this assessment agent to reach a final assessment.

In conclusion, the discussion above reveals that combining assessor agents, Fuzzy sets theory and negotiation in an assessment tool is an idea that deserves investigation.

**The evaluation modeling**

In this section, we present our two-stage evaluation approach, from the Fuzzy logic modeling of the individual skill assessment to the negotiation process taking place between assessor agents and mediated by the aggregate agent (see figure. 1). Our approach follows three steps:

**Step1: Skills identification step**

In order to better identify and to characterize the required skills within any of the domains of competence, we propose to follow the knowledge classification classically used in metacognition theory (Schraw & Moshman, 1995), where three types of knowledge are defined: *Procedural knowledge*, *Declarative knowledge* and *Conditional knowledge*.

![Diagram](image.png)

*Figure 1. General overview of the system*
Using these types of knowledge analysis, the main competences of the domain are first identified and are later divided in sub competences using a goal oriented analysis (see figure 2), until reaching the actions to be performed by the learner via the learning system interfaces. For example in the domain of Word-based mathematical problem solving, four main skills are identified: 1-understanding the problem, 2-making the plan of resolution, 3-executing the plan and 4-reviewing the solution. These competencies are considered as main skills that student must master to solve a Word-based mathematical problem. Then other skills are derived from the main ones until obtaining actions, reflecting identified skills, that student can carry out via the Computer based learning system.

**Figure 2.** Illustration of a competency model for word based mathematical problem solving

---

**Step 2: The step of skill evaluation**

In our evaluation approach, each skill within a domain of competence is represented by one or more actions that learner can perform, each action’s difficulty being qualitatively measured. We have defined according to Problem based learning experts, three levels of difficulty: (1) Actions with high level of difficulty, (2) Actions with average level of difficulty and (3) Actions with weak level of difficulty. Moreover, the scale used for the classification of learners’ competency control is inspired by the traditional classification of the performances employed by the assessors: (1) insufficient, (2) acceptable and (3) satisfactory (Merrill, 1983). For example, in the domain of Word based mathematical problem solving, Polya (1945) has identified four competences. In our system, each of them will be evaluated by an assessor agent through the evaluation of the learner’s performed actions (e.g., the competence called understanding the problem is characterized by many actions as shown in Table 1). To produce individual action evaluation, a fuzzy model is defined.

<table>
<thead>
<tr>
<th>Action</th>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifying important words in the problem</td>
<td>High</td>
</tr>
<tr>
<td>Identifying what is known in the problem</td>
<td>Average</td>
</tr>
<tr>
<td>Identifying what is requested in the problem</td>
<td>Average</td>
</tr>
<tr>
<td>Defining the position of the missing part</td>
<td>Weak</td>
</tr>
</tbody>
</table>

The fuzzy model represents the expert’s knowledge in linguistic form and includes the characteristics of the learner in the form of a set of fuzzy systems thus allowing an evaluation similar to that of an expert. A Fuzzy set is characterized by a membership function $f: E \rightarrow [0,1]$, which positions the members of the universe of discourse E in the units interval [0,1] (Zadeh, 1965). Value 1 (resp. 0) means that the member is (resp. not) included in the given set. The values between 0 and 1 characterize the fuzzy members. In our case, the universe of discourse E corresponds to the percentage of the actions performed correctly (classified by type and weighted by level of difficulty), and is divided in 11 elements {0,10,20,30,40,50,60,70,80,90,100}. For the transformation phase of quantitative variables into Fuzzy logical variables (i.e., Fuzzification process), we have defined four variables for each category of action (i.e., level of difficulty):

- **Three input variables** $VE_n (n \in \{1,2,3\})$ (one for each level of difficulty) and
- **One output variable** $VS$ representing the learner’s qualitative level of knowledge or mastership of one skill {Very weak, Weak, Insufficient, Average, Good, Very good, Excellent}. 

---

**Table 1. Illustration of actions to be performed**
For each one of these variables a membership function is defined in collaboration with experts in education (see figure 3 example). All assessor agents will use inference rules, based on these membership functions, to position the learner’s level on the classification scale. We have established 27 inference rules (3 levels of difficulty and 3 classification values for each level) corresponding to the various evaluations related to the three levels of difficulty (see figure 4 example).

Once computed, this fuzzy information must be converted into a real estimated value (i.e., defuzzified) which will represent the judgment of the assessor agent. The Defuzzification method produces a VS value as a number rounded to the nearest integer between 1 and 7 (1,2,3,4,5,6,7) corresponding respectively to the 7 evaluation levels (Very weak, Weak, Insufficient, Average, Good, Very good, Excellent) . At the end of the learning test session, each assessor agent has estimated the learner’s mastership level in “its” skill. This first stage evaluation will contribute to the overall evaluation of the second stage directed by the aggregate agent.

Step 3: The overall evaluation step

Generally, tutors set a threshold score for tests to be succeeded, for example in classrooms this score is fixed at 10 (i.e., half of full score). In our approach, we propose that each skill is considered mastered if the learner obtains a score equal or upper the value 5 (i.e., VS = 5, GOOD). However, we tolerate compensation between different skills with a minimum threshold score for each skill; unlike other domain of knowledge such as medicine where compensation is not accepted (Frank et al., 2010). In skills diagnosis, two classes of models have been commonly used: conjunctive models and compensatory models. When a domain of knowledge involves multiple skills and when the low mastery of a single one of them is sufficient for failing this domain knowledge, the model is considered part of the conjunctive class, meaning that all skills are necessary. Conversely, if a strong mastery of some skills is sufficient to succeed a test, it will be considered part of the compensatory class of models (Desmarais et al., 2012; Roussos et al., 2007). Our approach combines both model classes, by defining a minimum threshold degree of mastership of each skill and allowing compensation between skills by means of bonuses attribution for high level skill mastery.

At the reception of all evaluations (provided by assessor agents), the aggregate agent analyzes individually each one of them and computes the average score if necessary. Three decisions can be taken: the learner passes or fails the test, or the learner assessment needs negotiation. Analysis of the aggregate agent follows the algorithm shown in Figure 5.
Figure 5. The overall evaluation decision

Bonus marks benefits can improve students' attitude or effort, completion of work on time and/or neatness of finished work. In our approach, we propose to award bonuses to encourage the learner according to his skills performance. Obtaining a bonus allows to decrease the test validation score for a learner (initially equal to 5) of value "vb1" (i.e., value to be defined by domain of competence experts) for the first bonus and of "vb2" for the second bonus, this score diminution is used in the negotiation process to help learners whose global score is close to the validation score. For example, if a learner’s averaged score is equal to 4.66 (i.e., lower than the preset validation score, 4.66<5) and an assessor agent awarded him bonus1 and bonus2, the new validation score to be proposed by this assessor agent during the negotiation process for this learner is 5-(vb1+vb2).

**Computation details on how bonus marks guide the negotiation**

In our approach, three kinds of bonuses are awarded by agents according to student’s performance:

**Bonus 1**: each assessor agent decides on the attribution of bonus1 depending on the score obtained in the skill performance.

- **IF (skill's score ≥ 6)** then the assessor agent allots bonus1 to the learner (reduction in the validation score of a value of vb1, the agents accepts to validate the learner’s test with a score of 5-vb1).
- **IF (skill's score ∈ [5, 6])** then he allots half of bonus1 to the learner (reduction in the validation score of a value of (vb1)/2).
- **IF (skill's score < 5)** then he don’t allot bonus1 to the learner.

**Bonus 2**: is allotted by the assessor agent if the learner performs correctly at least 80% of actions with high level of difficulty, otherwise the learner does not receive this bonus. Obtaining bonus2 allows to decrease the global validation score of a value of vb2 (i.e., the assessor agent accepts to validate the learner’s test with a score of 5-vb2).

**Bonus 3**: the aggregate agent can allot one third bonus which consists in improving the learner’s score by a value of vb3 if this latter succeeded in at least 2/3 of all skills (e.g., if the learner’s averaged score of all skills is 4.66, the aggregate agent proposes to validate the learner’s test with a new score equal to 4.66+vb3).

Once bonuses are calculated, each assessor agent builds its satisfaction table and assigns a satisfaction degree to each validation score that he can propose to the aggregate agent during the negotiations (i.e., the satisfaction degree is used by the assessor agent as an indicator to propose or not a new validation score). This satisfaction table will be used in the process of negotiation; the assessor agent begins by proposing the first validation value in its table and continues until an agreement is reached or the actual degree of satisfaction is null (i.e., the assessor agent rejects the negotiation). An example of four satisfaction tables corresponding to four different cases of bonuses attribution is presented in table 2 (with vb1 = 0.34 and vb2 = 0.17) where no bonuses are attributed in case (a), only bonus2 is given in case (b), ½ bonus1 + bonus2 are attributed in case (c) and both of bonuses 1 and 2 are awarded in case (d).

For example in case (c), the first validation value is set to 4.83 (i.e., 4.83 = 5-0.17) after attribution of ½ bonus1, the degree of satisfaction corresponding to this value is 2 because there is another validation value: 4.66 (i.e., 4.66 = 4.83-0.17) computed after awarding of bonus2. The degree of satisfaction of this second validation value is 1 since there are no more proposals. The last value 4.5 has a null degree of satisfaction; this means that if a proposal of test
validation score is equal or less than 4.50, the assessor agent will reject this proposal. Note that the higher the degree of satisfaction is, the higher is the belief of the assessor agent to propose the corresponding validation value.

---

### Table 2. Four examples of satisfaction tables’ construction

<table>
<thead>
<tr>
<th>Case</th>
<th>Bonus</th>
<th>Validation Value</th>
<th>Degree of satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>5.00</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>4.83</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>4.83</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>4.66</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>4.83</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.66</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>4.50</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>4.83</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4.66</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.50</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>4.33</td>
<td>0</td>
</tr>
</tbody>
</table>

---

**From evaluation marks to agents negotiation**

The implementation of the negotiation needs the definition of a communication protocol between agents and the development of private strategies for each agent (Jennings et al., 2001). We present here the negotiation protocol used in our evaluation system and the rules employed by agents to reach a compromise. The objective of the protocol is to define the messages exchanged by the agents. The negotiation protocol that we propose is characterized by a succession of messages exchanged between an initiator (aggregate agent) and participants (assessor agents) as in Contract Net Protocol (Smith, 1980). We introduced a new idea in the negotiation process which consists in making a promise of concession by the aggregate agent if the assessor agent agrees to make a concession in turn (i.e., concession = one bonus consumption).

---

**Figure 6. Exchanging messages between the aggregate agent and the assessor agents**
Figure 6 describes the conversation between the aggregate agent and assessor agents. The aggregate agent starts by sending a CallForProposal message to the assessor agents, the parameter “g” indicates the group of concerned agents. All assessor agents reply by sending their proposals. The negotiation continues until aggregate agent sends an accept message to at least 2/3 of assessor agents (i.e., the test succeeds with a score equal to the last proposal done by assessor agents) or if the aggregate agent removes more than half of assessor agents from the negotiation process (i.e., the learner’s test failed).

Experiments and results

In order to verify the effectiveness of our assessment approach and to validate its performance, we have chosen the domain of competence of plant protection where a simulation crop pest control training environment is already developed to improve training of future phytosanitary agents. In this learning system, the learners’ behaviors have to be observed and analyzed in term of expected skills. However the systematic presence of a human expert cannot be ensured requiring some Computer-assisted assessment. The experts of this domain of competence have identified three main skills:

- **Control Skill**: as the ability of the learner to identify and to exploit several indicators (for ex. estimating a proliferation risk requires knowing current meteorological data, past data on pests’ location ...).
- **Intervention skill**: representing the achievement of one or several tasks correctly sequenced and accomplished at the right time.
- **Assessment skill**: the ability of the learner to assess a situation or a resources requirement considering scenario’s information and to act accordingly. It thus precedes one or several actions of intervention.

Following our two-stage evaluation process, we have developed three assessor agents and an aggregate agent. All agents are implemented in Jade (Java Agent Development Environment). Figure 7 displays messages exchanged between agents during the negotiation process and the final result is announced by the aggregate agent. Thresholds values (minimum test success score, maximum test failure score, bonus1, bonus2 and bonus3 values) were established by plant protection experts. The experiments we have conducted involved fifty students training in plant protection - zoophytiatry. These students were divided randomly into two groups: A (experiment group) and B (control group) of twenty five students each. To verify the effectiveness of our assessment system and to validate its performance, we conducted three experiments. Microsoft Excel was used for all data manipulation and statistical tests.

![Figure 7. The screenshot of the overall evaluation via negotiation between the assessors](image-url)
First experiment

The objective of this experiment is to test the effectiveness of the evaluation system, precisely, the first stage evaluation involving only assessor agents. To this end, students in group A were assessed during two learning test sessions by the system (i.e., our three assessor agents), and after adaptation of the same test scenario on paper, the responses of these same students were again assessed by an expert of the domain taking into account the number, type and difficulty of the actions carried out correctly. In order to avoid bias between the system’s evaluation and that of the expert one, we intentionally conducted two learning test sessions. We adopt the Pearson correlation to compare the correlations between scores given by the system (assessor agents’ scores) and scores of the expert for each type of skill and for both sessions (i.e., all scores are between 1 and 7, see section 3.2). Table 3 illustrates this experiment’s results.

<table>
<thead>
<tr>
<th>Type of competence</th>
<th>Session 1 Correlation</th>
<th>Session 1 Critical value</th>
<th>Session 2 Correlation</th>
<th>Session 2 Critical value</th>
<th>Level of Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control skill</td>
<td>0.7618</td>
<td>0.396</td>
<td>0.8991</td>
<td>0.396</td>
<td>0.05</td>
</tr>
<tr>
<td>Intervention skill</td>
<td>0.8111</td>
<td>0.8326</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment skill</td>
<td>0.8101</td>
<td>0.7144</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion: This first experiment intended to examine the effectiveness of our assessment system. The best way to achieve this is to compare the scores assigned by the system to those given by the domain expert during two rounds. Significant correlation (correlation is greater than the critical value for the two sessions) between the two evaluations in this experiment constitutes a preliminary validation of the effectiveness of our assessment system. We are however aware of the limitations of our validation since our approach is conducted in only one simulation learning system. Also, the results validate the claim regarding the importance of embedding Fuzzy sets in assessment process. By providing actions with several levels of difficulty, the assessment system is able to accurately point the learner’s level of knowledge with a certain granularity for each competence of the domain. This has some influence in the training system software requirements.

Furthermore, dedicating specialized agent for assessing each skill individually has several advantages: The diversity of agents enables a distributed assessment and ease of expandability of assessed skills. Heterogeneous agents can thus evaluate specific information (learner’s answers versus predefined correct answers, simulation results, users' interactions with the training system etc.). As such, the use of agents allows defining independently each skill assessment method (different criteria, parameters and algorithms).

Second experiment

In order to test the performance of students evaluated by the system, we have organized two learning test sessions where students in Group A are evaluated by the system and receive feedback at the end of the first learning test session allowing them to revise their behaviors. For the same test scenario, students in Group B are evaluated by the domain expert and receive no explanation on their behaviors. The comparison between behaviors during the first session and those of the second session, for both groups, allows to rule on the performance of the students. We propose a paired t-test analysis to verify the improvement of behaviors in each group. Table 4 represents the analysis of the performances for the group A and the group B.

<table>
<thead>
<tr>
<th>Type of competence</th>
<th>Group A</th>
<th>Group B</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t-value</td>
<td>p-value</td>
<td>Session 1 mean</td>
<td>Session 2 mean</td>
<td>t-value</td>
</tr>
<tr>
<td>Control skill</td>
<td>2.6134</td>
<td>0.0152</td>
<td>3.92</td>
<td>4.40</td>
<td>1.6164</td>
</tr>
<tr>
<td>Intervention skill</td>
<td>2.3842</td>
<td>0.0254</td>
<td>3.52</td>
<td>4.12</td>
<td>1.7321</td>
</tr>
<tr>
<td>Assessment skill</td>
<td>2.5981</td>
<td>0.0158</td>
<td>3.32</td>
<td>3.92</td>
<td>2.00</td>
</tr>
</tbody>
</table>
Discussion: Regarding the performance of students evaluated by the system, we have conducted two learning test sessions. Comparison of students’ scores obtained during the first and second learning test session and compared to those of the first session indicates whether or not learners have improved their behaviors. Improvement of behavior between the two sessions confirms the accuracy of assessment and the quality of feedback. As all p-values (group A’s experiment) don’t exceed the level of significance \( \alpha \) (with \( \alpha = 0.05 \)) we conclude that the difference between the two groups was significant. Analytical results indicate that students of group A had acquired more knowledge than students of group B since the means of scores for two types of competence (control and intervention) increased significantly from 3.92 and 3.52 to respectively 4.40 and 4.12. However, for the assessment skill, means of scores gained a slight improvement from 3.32 to 3.92; we believe this is due to the assessing complexity of some situations in the simulation learning environment. Furthermore, means of scores of students in group B remained constant except for assessment skill where scores have risen moderately. All these results point to the conclusion that learners learned from the feedback generated by the assessment system.

Most students of group A confirm that the system helped them to review and improve their behaviors. They also felt that the flexibility of the assessment through bonuses’ awarding and negotiation is a motivating factor that prevents loss of motivation and give up of the learning system besides improving the overall behavior of the learner. However, some would think that if students are involved in the evaluation they could argue their choice of actions.

In summary, assessing learners differently than existing approaches, which assessment frequently focus on the result regardless the learner’s skills involved to obtain this result, allows us to address the problem of subjective assessment and overcome some difficulties encountered in traditional measurement models. Our approach can support learners to build problem-solving skills, and to enhance their behaviors over learning test sessions through independent assessor agents.

### Third experiment

In order to measure the impact of the Negotiated assessment on student’s motivation to continue learning and not give up the learning system, we have organized six learning test sessions for both groups (2 sessions per day). Students of group A were assessed using both stages of the evaluation system (i.e., individual skills assessments and the negotiation mechanism) while students of group B were evaluated only by the first stage of our evaluation model without any possibility of negotiation between assessor agents (if the student’s score \( \geq 5 \) then the test is succeeded else it failed). Specifically, we focus on the participation rate and the dropout rate of both groups’ students having obtained consecutively during many sessions a score in the interval \([4.33, 5]\) (i.e., \( \text{min}_\text{sc} = 4.33 \), this value was fixed by the domain experts, it allows negotiation between assessor agents if the score of a learner in group A is \( > 4.33 \) and punishes a student of group B by a failure of test if his score is \( < 5 \)). Note that for this experiment, we do not consider the abandonment of learners with other scores. The students were informed that they could leave the learning test session at any time and give reasons for their abandonment by mean of the questionnaire provided at each session. Table 5 shows respectively for each group the number of learners having obtained successively for \( x \) times a score in the interval \([4.33, 5]\) (with \( x \in [2,6] \)), followed by the number of learners among them who have succeeded their learning test at this attempt (i.e., by mean of negotiation), followed by dropout rate of this attempt. All 50 students attended the session 1.

<table>
<thead>
<tr>
<th>Attempt</th>
<th>X = 2</th>
<th>X = 3</th>
<th>X = 4</th>
<th>X = 5</th>
<th>X = 6</th>
<th>Total of dropouts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>14/7/0</td>
<td>10/7/0</td>
<td>9/4/0</td>
<td>8/5/1</td>
<td>5/4/1</td>
<td>2</td>
</tr>
<tr>
<td>Group B</td>
<td>9/0/0</td>
<td>7/0/1</td>
<td>5/0/3</td>
<td>2/0/2</td>
<td>0/0/0</td>
<td>6</td>
</tr>
</tbody>
</table>

Discussion: We note that the dropout rate in group B was higher than that in group A for that category of learners although it is not very high (i.e., 6 Vs 2). In order to identify the reasons of this abandonment, we have analyzed the answers of targeted learners to the questionnaire provided for indicating the reasons of giving up the learning test session (see Table 6). In addition, the low dropout rate of group A’s learners is due to the high success rate of learners unlike learners of group B by dint of the flexibility of the negotiated assessment as shown in Table 5.
Table 6. Learners’ answers about giving up the learning test session

<table>
<thead>
<tr>
<th>Question: Why you no longer want to continue learning tests?</th>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>The test is difficult</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>The evaluation is subjective</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>The learning environment is not interesting</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>The simulation learning environment does not reflect the reality on the ground</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

All answers (except one) of group B’s dropouts (5 of 6) for this category of learners point to the conclusion that the evaluation is subjective, students report that they deserve to succeed their learning tests sessions at many times while the system considered they have failed, it is precisely for this reason that we chose this category of learners (close to success, but unfortunately due to a rigid common evaluation, the system decides that they failed their tests). While in group A, only one learner having forsaken the learning test session considered that the evaluation is subjective. This comparison points to the conclusion that the main cause of the abandonment of learners is due to the subjectivity of the adopted assessment mechanism, while the use of the Negotiated evaluation for learners of group A has contributed to the preservation of motivation to continue learning. In addition, we took no heed of dropouts of learners with different consecutive scores (not all belonging to the interval [4.33, 5]) so as not to bias the results of the experiment. Although we are aware of the reduced sample of this experiment, we belief that a large-scale experiment will confirm our claim.

Limitations of the present study

- The participation of students was voluntary and we do not claim that the results of the present study can be generalized.
- The improvement of learner’s behavior performance can be attributed to the use of the proposed evaluation model or to the extra practice by solving similar pest control issues. Even in the third experiment, the profit for education is that students preserve motivation for learning even if the experiment is based on a partial comparison between the experimental and the control groups. Undoubtedly the educational phenomena are multidimensional and we cannot control all the possible involved variables.

Conclusion

This study has presented a pragmatic approach of evaluation that allows giving a ruling on the efficiency of the learners in a context of Multi-agent simulation learning systems. This evaluation approach is to be integrated in Simulation-based learning systems, Game-based learning or any Computer assisted problem-solving based learning. Our evaluation model follows three steps: (1) Identification of relevant skills of the domain of knowledge, (2) Evaluation of the learner compared with these skills and (3) Evaluation of the ability of the learner to solve a problem of this domain of competence. In order to assess learners’ competences, we have adopted a strategy that involves a two-stage approach based on a collaborative evaluation system. On the first stage, a number of assessor agents are in charge of appreciating the learners’ knowledge compared with identified skills of the domain. The recourse to Fuzzy sets theory at this stage allows an evaluation similar to that of an expert. On the second stage, the Aggregate agent ensures an overall evaluation elaborated on the base of the individual evaluations of each assessor agent. Instead of considering an average threshold score which decides of the success and the failure of the learner’s learning test, we have preferred a Negotiated collaborative evaluation similar to academic evaluations which promote students according to their results along the learning period. Thus, in some situations, a negotiation process is initiated by the Aggregate agent where each assessor agent uses the learner’s results in its dedicated skill in order to negotiate the learner’s test success by mean of bonuses awarding. The result of this negotiation represents the final evaluation.

Experimental results indicate that this model provides an assessment similar to that of an expert and significantly improved learners’ performance. Furthermore, the Negotiated assessment part of the evaluation model seems promote motivation of learners as demonstrated in third experiment. Based on these results, we conclude that combining Fuzzy sets and agent negotiation has important merits.
Our evaluation system allows initially a skill estimate, this first stage informs us about the strengths and weaknesses of the learner, and thus we provide high precision recommendations to the learners. Therefore the quality of feedback will generate a positive impact on improving the learner’s skill performance. In the second time, the global evaluation enables us to conclude about the effectiveness of the learner to solve the problem. Thus, tutors can easily address learners’ weaknesses.

Finally, although the proposed approach of assessment has yielded promising results in promoting learning effectiveness and maintaining students’ motivation to continue learning, considerable work remains to be done, including the choice of adequate problem corresponding to the learner’s profile and based on his skills assessments. Also we’ll consider other learner’s abilities in the future like group managing, communication, etc in the context of collaborative problem solving.

References


An Exploration into Improving Examinees’ Acceptance of Participation in an Online Exam

I-Fan Liu, Ruey-Shin Chen* and Hao-Chun Lu
Department of Information Management, Oriental Institute of Technology, Taiwan // Department of Computer Science and Information Engineering, National Quemoy University, Taiwan // Department of Information Management, Fu Jen Catholic University, Taiwan // ifliuG@gmail.com // rueyshin_chen@nqu.edu.tw // bach0809@gmail.com

*Corresponding author
(Submitted October 19, 2013; Revised May 22, 2014; Accepted July 26, 2014)

ABSTRACT
With the rapid development of the Internet and information technology, the issues related to online exams have become the concern of an increasing number of researchers. At present, the biggest challenges for the integration of web communication technology into online exams are the ability to detect cheating behaviors during the exam, and the verification of students’ real identification. These problems can have an impact on the students’ perceived trust of an online exam, and further influence their participation acceptance. This study proposes a research model with notarization process (technical level) and online exam policy (institutional level) in order to improve examinees’ participation in online exams. There are 301 college students participating in this study. The results indicate that all the hypotheses in this model are supported, which implies that the monitoring and managing mechanisms proposed in this model can significantly raise examinees’ participation acceptance for online exams. Finally, we further discuss the implications of the research findings and provide some suggestions for future developments for online exams.

Keywords
Perceived fairness of notarisation process, Perceived soundness of online exam policy, Perceived trust, Participation acceptance

Introduction
Due the capacity of the Internet to transcend time and space, traditional paper and pencil exams have gradually been transformed from physical space to cyberspace. Online exams have been widely used in many institutions as a convenient alternative assessment tool for evaluating students’ learning outcome. However, in current educational systems, there is a considerable gap in the trust of online exams compared to traditional ones. It is difficult for examiners to monitor and credibly assess online exam systems as a result of the difficulties relating to the identification process of the students. The system should be capable of confirming whether the examinees are in fact the students specified by their identification. It is generally acknowledged that cheating in exams is not uncommon among students on college campuses (Bolin, 2004). Even though the use of technology can mitigate these unfavourable phenomena but cheating cannot be completely eradicated (Pain & Le Heron, 2003). Studies have uncovered increasingly diversified trends in cheating behaviours, and the fairness of exams is increasingly being called into question (Crown & Spiller, 1998). According to a study of online exams, 45% of students self-reported negative behaviours, such as lying, cheating, and defiance, during online exams (Russell & James, 2008), and 73.6% of the students in the sample believed that it is easier to cheat in an online exam than in a traditional exam (King, Guyette, & Piotrowski, 2009). Of these, one of the main reasons that the fairness of online exams is being called into question is the user authentication problem (Apampa, Wills, & Argles, 2010). For example, if a student with poor grades were to find a student with good grades to take an exam on his behalf, this could lead to benefits beyond the exam itself, such as higher a chance of success when applying to a more reputable graduate school. Once the fairness of exams has been violated, the trust in the system from the examinees decreases. Therefore, trust has been recognised as the most important factor in a successful participation of online exam website (Chang, Tseng, Chou, & Chen, 2011). Given the convenience of the Internet, governments and educators are thus challenged with the task of establishing fair environment and policy for online exam. As the need for online exams continues to grow, more complete user identification, including a face recording system for examinees, has been suggested for online exam websites (Chen, Hsieh, & Tsai, 2008). Furthermore, school administrations should also develop a sound policy to prevent cheating behaviours in online exams. In order to minimize cheating behaviours on online exams, we approached the issue from the perspective of two mechanisms: one at the technical level and the other at the
institutional level. Therefore, this study proposes a notarisation process (technical level) combined with the setting of an online exam policy (institutional level) to develop a more reliable and fairer exam environment. Moreover, we also propose a hypothetical model as our research framework to explore other factors whether affect examinees’ participation acceptance of online exams.

**Existing approaches to verify a user’s identity**

In the context of online exams, user authentication refers to the security procedure examinees use to establish a trust relationship with the system. It is similar to the checks on examinees’ identity (ID) cards that examiners perform before allowing them into a traditional exam room. Currently there are several approaches used to verify examinees’ real identities, such as face recognition, which is a method that can successfully and safely confirm personal identification (Chen et al., 2008). Other common approaches are listed below with brief descriptions:

**Username and password**

For this method, there is no need to purchase additional expensive equipment, and all that is required is a computer equipped with a mouse and keyboard. This is the most well-known and widely used authentication method. The main drawback of this method is that the password may be leaked and the true examinee may be impersonated by someone else.

**Smart cards**

Smart cards have been developed as the means of identifying users to replace the use of accounts and passwords. However, this method is impractical and limited as the hardware owned by examinees needs to identify the smart card, and it is still possible that the students may give their identity cards to others. For example, student A uses student B’s smart card to login in the system to take exam, a condition leading to the inability of the system to detect the real identity and resulting in the problem of cheating.

**Biometric authentications**

Biometric methods can be divided into active and passive types depending on how the user interacts with the system during the authentication process. Such as fingerprinting and face recognition, consist of the use of biometrics to identify users in order to not only resolve the user authentication problems but also offer a higher level of information safety and ensure users’ identities (Xiao & Yang, 2009). Generally speaking, fingerprinting is an active authentication style, which requires users to actively provide biometrics such as fingerprint information before they use the system. If a re-authentication request is needed during the exam, students have to use their fingers again for identification, a condition which may distract students’ attention from the exam. Face recognition, in contrast, is a passive authentication style. Users only need to use a webcam in front of their monitors, and the system will keep recording, identifying, and eliminating interferences to examinees during the exam. Fingerprints and face recognition provide much safer authentication processes than passwords. However, these types of biometric methods require continuous deliveries of re-authentication requests to users, which can interfere with the online exam process of the examinees. Also, a series of re-authentication processes requires large amounts of computation which can cause an additional load on the system (Stallkamp, Ekenel, & Stiefelhagen, 2007).

**The notarisation process of online exams**

In the domain of information security, a public key is subject to encryption and decryption technology that not only makes it harder to tamper with the original data but also ensures their authentication (Rivest et al., 1978). After this, a third party such as a certificate authority is needed to notarise the user name and create a connection between the public key in a process known as a concept of notarisation (Ansper, Buldas, Roos, & Willemsen, 2001; Lekkas & Gritzalis, 2004). To verify the real identity of the examinees and at the same time avoid interference to examinees
and additional system load from continuous re-authentication, this study offers a mechanism that combines face notarisation and face recording in online exams. The purpose is to create a fair online exam environment through the notarisation process.

This study provides a brief sketch of the notarisation process in an online exam (see Figure 1). During course registration, the exam system captures one face picture randomly from a webcam and saves it in the student’s basic record. In the procedure 1, after a student registers in the online exam system and uploads a photograph of his or her face, the student arrives at the exam and is confirmed by the proctor or TA (as shown in Figure 2). In procedure 2, during the exam, the system randomly takes three photographs of the student’s face and records these photographs with the student’s image test report. In procedure 3, a teacher, staff member, student or TA can review the report following the online exam. In a normal situation, the recorded photographs will represent the registered student. This study implements a notarisation process and proposes a mechanism for cumulative face notarisation to explore whether the findings indicate that examinees will have a greater trust in modern online exams.

**Research model**

**Theoretical development**

The technology acceptance model (TAM) suggested by Davis (1986) explores the acceptance of new information technology by users. In their model, two determinants have been found to have a significant impact on the intention to use new technology: perceived usefulness and perceived ease of use. Both of these two variables are affected by external variables (Davis & Venkatesh, 1996). Both of the above variables impact on users’ acceptance of information technology. In the causal relationship of the entire learning process, learning is the cause, and assessment is the result. Regarding learning, when students feel that their online learning experience can be
improved by virtue of the usefulness and ease of use brought by learning materials offered by information technology, they are more willing to use the system in the future (Pituch & Lee, 2006). Several studies have found that, in addition to the learning content, information quality and network quality also influence learners’ satisfaction of using the system (Piccoli, Ahmad, & Ives, 2001; Webster & Hackley, 1997). Other research findings also indicate that the higher the quality and reliability of information technology, the better the learning outcomes (Piccoli et al., 2001; Webster & Hackley, 1997). This further suggests that learners’ acceptance of the integration of IT into learning is high.

However, in terms of assessment, it is participation acceptance, rather than technology acceptance, that is the key concern of our study. Factors such as perceived fairness of notarisation process, perceived soundness of online exam policy and perceived trust are related to students’ participation acceptance in online exams. Hence, in the following subsection of literature review, we will describe how these factors impact on participation.

**Perceived fairness of notarisation process**

Not only is online assessment not limited to time and place, as is also the case with online learning systems, but it can also improve the process of traditional assessment method insofar as online exams can save time, offer immediate feedback, reduce resources, keep detailed records and be conveniently used (McCormack & Jones, 1997). However, the fairness of the exam is a very important issue because of the typical notarisation issues that plague online exams. For example, if an online exam cannot identify users, cheating will happen more often (Apampa, Wills, & Argles, 2009). Similar online learning systems, online assessment offers practice and positive learning outcomes (Buzzetto-More & Alade, 2006), although fairness of exam outcomes should be ensured to raise examinees’ belief in the trust of as well as their participation acceptance of online exams (Asha, 2008; Doukas & Andreatos, 2006). On the other hand, some studies have noticed that the fairness of online testing has an impact on examinees’ participation and have proposed solutions. One such general solution is the use of users’ accounts and passwords for identification. Although this is an easy and general method, it is frequently questioned, such as when leaked passwords immediately jeopardize the fairness of the online exam (Apampa et al., 2009). The outcomes of the exam will naturally be questioned if fairness and justice cannot be assured, leading to the loss of meaning of the exam and its impact on learners’ participation acceptance for online testing in the future.

In this study, the purpose of the face notarisation coupled with recording system was to verify the identity of each student and prevent cheating behaviours (e.g., impersonation) in online exams in order to achieve notarised fairness. Online exams will only be fair if the system can effectively identify the real identity of examinees. The construct Perceived Fairness of Notarisation Process mainly measures how examinees perceive the fairness of the notarisation process of online exams. And we also explore its impact on perceived trust and participation acceptance. We make the following two hypotheses.

H$_{1.1}$. The Perceived Fairness of Notarisation Process will positively influence the examinee’s Perceived Trust of online exams.

H$_{1.2}$. The Perceived Fairness of Notarisation Process will positively influence the examinee’s Participation Acceptance.

**Perceived soundness of online exam policy**

In a typical exam setting, classroom regulations focus primarily on limiting students’ behaviour: students may not talk or help to each other during the test process. Bugbee (1996) argues that under the exam principles of development, usage, and explanation, computer-based exams have the same efficacy as paper-based ones. Some studies, however, indicate that students tend to violate exam rules in computer-based exam (Buzzetto-More & Alade, 2006). For example, although students comply with exam regulations and act appropriately when taking a traditional exam monitored by instructors, students may use uncharacteristic strategies to obtain better grades on a web-based exam, such as whispering to others in the exam (Apampa et al., 2009). This situation may give rise to feelings of guilt and a sense of violating one’s integrity. Aside from putting forward a notarisation process to prevent cheating behaviours in online exams, this study also stresses the need to establish a complete online exam policy (Schroeder, 2013; Wilson, 2004) that increase examinees’ trust and participation acceptance in online exams. Online exam policy is defined here as all strategies and regulations used to manage online exams, including punishments for cheating,
credit acknowledgement, question type design, and so on. For example, computerised adaptive testing gives examinees different question items according to their abilities (van der Linden & Glas, 2000), and other studies have proposed methods like randomised items in order to prevent cheating behaviours (Marks & Cronje, 2008; Pain & Le Heron, 2003). These policies above are related to acceptance of participation and the trust of online exams. The construct Perceived Soundness of Online Exam Policy mainly measures how examinees perceive the policy soundness of online exam. In this way, the perceived trust of online exam and examinees’ acceptance to participate it are examined through the following hypotheses related to perceived soundness of online exam policy.

H_2-1. The Perceived Soundness of Online Exam Policy will positively influence the examinee’s Perceived Trust of online exams.

H_2-2. The Perceived Soundness of Online Exam Policy will positively influence examinee’s Participation Acceptance.

**Perceived trust and participation acceptance**

A reliable system will make an impact on learners’ intention to participate in e-learning (Ong, Lai, & Wang, 2004). Many studies have suggested that the security of information systems (such as systems of e-commerce, e-learning, and e-assessment) has an influence on users’ participation (Chen et al., 2008; Ong et al., 2004). That is, when information system quality is reliable, stable, and creditable, users will have higher intention to participate.

Students’ learning performance is usually based on the score from the assessment. The evaluations of students’ performance are influenced by traditional principles of assessment, such as fairness, trust, efficacy and comparability (Messick, 1994). During the process of a traditional test, examiners are able to monitor examinees’ behaviours to detect cheating. Perceived trust offered by computer-based or web-based exam will make an impact on students’ participation for these types of exams (Chang et al., 2011; Newhouse, 2011). Regarding to the Participation, it can be an indicator of students’ acceptance of online exams. If the system cannot provide a teacher-assessment environment, the fairness of the online exam will be questioned and challenged by examinees. If students do not trust the exam system, their acceptance to participate in exams will be affected even though teachers have the right to decide the exam format (Chang et al., 2012; Jaillet, 2009; Smith, 2007).

Web innovativeness affects participation intentions (Fang, Shao, & Lan, 2009). In this study, the notarisation process with face recording (technical level) and online exam policy (institutional level) is an innovative implementation that may act as an incentive to attract students to use the online exam system and may influence the acceptance to participate. The construct Perceived Trust measures overall how much perceive trustworthy the examinees have in the online exams, including the perception from technical and institutional levels. The construct Participation Acceptance is used to measure the overall acceptance of participation examinees feel towards online exams. To examine Perceived Trust and Participation Acceptance, the following hypotheses are proposed. Based on the above literature, this study explores whether examinees’ participation of an online exam in our research model to discuss and illustrate the relationships between all factors that influence the examinees’ participation in online exams (see Figure 3).

H_3. The Perceived Trust of examinee will positively influence the examinee’s Participation Acceptance.
Methodology

Instruments

The research model proposed in this study had its question items based on four constructs. Some items of the constructs (Perceived Fairness of Notarisation Process, Perceived Soundness of Online Exam Policy) were developed in this study. The items in Perceived Trust and Participation Acceptance constructs were adapted from previously validated instruments for use in an online exam context (Chen et al., 2008; Ong, Lai, & Wang, 2004; Davis & Venkatesh, 1996). In total, 20 question items comprised the preliminary questionnaire. A five-point Likert type scale was adopted to measure the extent of examinees’ feelings towards the four constructs from 1 point (strongly disagree) to 5 points (strongly agree). For each construct, sample constructs were compared to identify significant differences. It was hypothesized that significant differences in the participation Acceptance of the online exam would exist between different constructs (Perceived Fairness of Notarisation Process, Perceived Trust, and Perceived Soundness of Online Exam Policy). We collected 128 samples from the pre-test. With the Cronbach’s alpha value greater than 0.8, we found the questionnaire to be reliable in the pre-test.

Online exam context and participants

A total of 325 college students from northern Taiwan attending a class of English literature reading given at a centre for General Education took part in the study. This class was split into two semesters. Before the semester ended, the students' learning outcomes were assessed through online exams. The main type of question items in this study was multiple choice questions. The instructors needed to first upload pre-set question items to a database and also to define the difficulty of each item according to the question content, with difficulty levels ranging from difficult, medium, and easy. For example, an instructor might wish to set an exam level that had 20% with difficult, 50% with medium and 30% with easy questions. The instructors only needed to adjust the difficulty settings of exams to the aforementioned proportions, and the system would then randomise the exam items for each examinee. Therefore, while examinees were examined under the same level of difficulty, they received different exam items. In addition, the exam took place in the computer centre of the school where examinee seating was determined randomly on the spot so it was impossible to know beforehand where the examinees would be sitting.

The study used a face recording system, a webcam embedded in a computer monitor (see Figure 4), at the beginning of course registration and added a notarisation process to verify the identity of each student. At the beginning of a class, a teacher or teaching assistant confirmed students’ identification using the students’ photos (see Figure 5). If students had registered and uploaded a photograph, the system showed a photo confirmation date and notes who confirmed the photo. Then, the registered photos were compared with the photos captured by the online exam process (see Figure 6). After the exam, the students were asked to complete a questionnaire. After the removal of invalid samples, a total of 301 valid ones remained. The response rate was 92.6%.
Data analysis

This study adopted the maximum likelihood method to estimate the model’s parameters. According to the subject-to-variable ratio, the ratio should be no lower than 5 (Bryant & Yarnold, 1995). Structural equation modelling (SEM) is used in this study in order to estimate the measurement and structural model so as to determine whether the study’s assumptions are a good model fit after analysis and modification (Ngai, Poon, & Chan, 2007).

To examine this model, the first step is to determine the construct validity, which involves the verification of the hypotheses of this study. Factor analysis and path analysis are the basic tools used to verify the construct validity of a model. Table 1 lists the findings from the exploratory factor analysis (EFA). All Items are good fit into 4 factors.

Table 1. Results of exploratory factor analysis

<table>
<thead>
<tr>
<th>Items</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Trust 1</td>
<td>.771</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Trust 2</td>
<td>.744</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Trust 3</td>
<td>.687</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Trust 4</td>
<td>.598</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Trust 5</td>
<td>.578</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation Acceptance 1</td>
<td></td>
<td>.785</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation Acceptance 2</td>
<td></td>
<td>.697</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation Acceptance 3</td>
<td></td>
<td>.643</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation Acceptance 4</td>
<td></td>
<td>.639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation Acceptance 5</td>
<td></td>
<td>.617</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Fairness of Notarisation Process 1</td>
<td></td>
<td></td>
<td>.750</td>
<td></td>
</tr>
<tr>
<td>Perceived Fairness of Notarisation Process 2</td>
<td></td>
<td></td>
<td>.709</td>
<td></td>
</tr>
<tr>
<td>Perceived Fairness of Notarisation Process 3</td>
<td></td>
<td></td>
<td>.678</td>
<td></td>
</tr>
<tr>
<td>Perceived Fairness of Notarisation Process 4</td>
<td></td>
<td></td>
<td>.566</td>
<td></td>
</tr>
<tr>
<td>Perceived Soundness of Online Exam Policy 1</td>
<td></td>
<td></td>
<td></td>
<td>.815</td>
</tr>
<tr>
<td>Perceived Soundness of Online Exam Policy 2</td>
<td></td>
<td></td>
<td></td>
<td>.793</td>
</tr>
<tr>
<td>Perceived Soundness of Online Exam Policy 3</td>
<td></td>
<td></td>
<td></td>
<td>.680</td>
</tr>
<tr>
<td>Perceived Soundness of Online Exam Policy 4</td>
<td></td>
<td></td>
<td></td>
<td>.584</td>
</tr>
</tbody>
</table>
Initially, the questionnaire contained 20 items, but two poorly performing items were deleted during exploratory factor analysis (EFA) to make the model more stable. Thus, the final version of the questionnaire contained 18 items (see Appendix A). According to Nunnally and Bernstein (1994), Cronbach’s alpha can be deemed reliable if its value is at least 0.7. The value of Cronbach’s alpha for the four constructs in this research was greater than 0.87.

The average variance extracted, which is used to measure the discriminant validity of each construct, is only acceptable when it is greater than 0.5 (Fornell & Larcker, 1981). The average variance extracted was generally higher than 0.70. Additionally, the KMO test was 0.96, which indicates that the questionnaire was well-designed in this study. The reliability and validity of this questionnaire are both acceptable.

**Model testing results**

Many indices can be used to assess the fit of a model (Schumacker & Lomax, 1996). This study uses indices suggested by Hoyle and Panter (1995). In general, the closer the observed data is to the research model, the better the fit of the model, and the easier it will be to satisfy the thresholds of index values. As shown in the presentation of this research model in Figure 7, and the criteria for the evaluation of the model: (1) $\chi^2$/d.f. = 2.41 (< 3.0); (2) the Goodness-of-fit index (GFI) = 0.95 (> 0.9); (3) the Normed fit index (NFI) = 0.97 (> 0.8); (4) the Relative fit index (RFI) = 0.96 (> 0.9); (5) the Incremental fit index (IFI) = 0.98 (> 0.9); (6) Root mean square residual (RMR) = 0.059 (which is close to the recommended value less than 0.05); (7) and the Root mean square error of approximation (RMSEA) = 0.068 (< 0.08). Most index values comply with the threshold values, and the estimates of the regression weights are significant at $p < 0.05$ (as shown in Table 2).

![Figure 7. The results of research model](image)

### Table 2. Standardized regression weights

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Independent variable</th>
<th>Dependent variable</th>
<th>Estimate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1.1</td>
<td>Perceived Fairness of Notarisation Process</td>
<td>Perceived Trust</td>
<td>0.20</td>
<td>***</td>
</tr>
<tr>
<td>H1.2</td>
<td>Perceived Fairness of Notarisation Process</td>
<td>Participation Acceptance</td>
<td>0.22</td>
<td>0.001**</td>
</tr>
<tr>
<td>H2.1</td>
<td>Perceived Soundness of Online Exam Policy</td>
<td>Perceived Trust</td>
<td>0.42</td>
<td>***</td>
</tr>
<tr>
<td>H2.2</td>
<td>Perceived Soundness of Online Exam Policy</td>
<td>Participation Acceptance</td>
<td>0.14</td>
<td>0.047*</td>
</tr>
<tr>
<td>H3</td>
<td>Perceived Trust</td>
<td>Participation Acceptance</td>
<td>0.50</td>
<td>***</td>
</tr>
</tbody>
</table>

Figure 7 shows the causal relationship between the constructs and the standardised path coefficients. In the results of the $H_{1.1}$, face notarisation and recording technology increase the fairness of the online exam. This factor of Perceived Fairness of Notarisation Process significantly ($^*^*^* \ p < 0.001$) affects the Perceived Trust of the online exam.

From the results of the $H_{1.2}$, the factor Perceived Fairness of Notarisation Process significantly ($^*^* \ p = 0.001$) affects the examinees’ Participation Acceptance for the online exam. The results from $H_{2.1}$ also reveal that the antecedent variable of Participation Acceptance, Perceived Trust, is significantly ($^*^*^* \ p < 0.001$) affected by the Perceived Soundness of Online Exam Policy.
The results from $H_{3-2}$, it can be seen that the Perceived Soundness of Online Exam Policy significantly ($p < 0.05$) affect Participation Acceptance. Finally, the results from $H_3$ indicate that the antecedent variable, Perceived Trust does significantly affect ($p < 0.001$) the Participation Acceptance. Hence, $H_{1-1}$, $H_{1-2}$, $H_{2-1}$, $H_{3-2}$, and $H_3$ are all supported. Further, the explained variances include Perceived Trust ($R^2 = 0.52$) and Participation Acceptance ($R^2 = 0.46$).

As depicted in Figure 7, there is a strong relationship between Perceived Soundness of Online Exam Policy and Perceived Trust ($\beta = 0.42, p < 0.001$), as well as Perceived Trust and Participation Acceptance ($\beta = 0.50, p < 0.001$). This suggests that the completeness of the online exam policy has an impact both on students’ perceived trust of the exam process, as well as directly and indirectly on students’ participation acceptance for the online exam. In addition, although the relationship between the Perceived Fairness of Notarisation Process and Perceived Trust ($H_{1-1}$, $\beta = 0.20, p < 0.001$) is less strong than that between Perceived Soundness of Online Exam Policy and Perceived Trust ($H_{3-1}$, $\beta = 0.42, p < 0.001$), the Perceived Fairness of Notarisation Process actually has a statistically significant impact on the Perceived Trust of the exam process, and directly and indirectly on the Participation Acceptance of the online exam. This finding importantly confirms that students will more likely trust and participate in future online exams.

Discussions and conclusions

This study discovered that the most significant determinant affecting examinees’ Participation Acceptance for the online exam was Perceived Trust. This implies that the more reliable and credible examinees feel about the online exam, the more acceptance they will be to participate in it. This finding is consistent with prior studies (Chang et al., 2011; Newhouse, 2011). After all, an online exam via a webcam and the Internet is relatively stranger and more unfamiliar to examinees compared to a paper-based test in the traditional classroom. For most examinees, it is a whole new experience to take this type of exam. In Taiwan, during paper-based tests, one or two examiners typically monitor the exam in order to prevent cheating behaviours. Examinees can submit their examination papers after answering all the questions, and they don’t have worry about possible system problems associated with online exams, such as system crashes, network disconnections, or unexpected situations due to unfamiliarity with the operation of online exam system. Even though examinees may want to cheat during an exam, they have to bear the risk of being punished with a zero score or possible expulsion, which is an effective deterrent for most. This mechanism of with monitoring and punishment is believed to be essential for fairness and test trust crucially depends on it. As such, the same mechanism should also necessarily be established for online exams.

In order to increase examinees’ perceived trust of an online exam, the study aimed to develop an online exam environment from the two perspectives of technical level and institutional level. Regarding the former perspective, this study found that when examinees felt the notarisation process was fair, they perceived the exam to be more reliable. In order to prevent the problem of fake ID due to the inability to verify users’ authentic identity (Apampa, 2010), we combined face notarisation with face recording to increase the system’s capability of verifying real identity of each examinee to avoid fake ID or other cheating behaviours. In traditional exams, examiners have to monitor all examinees in the classroom to detect other cheating behaviours, such as using cheat sheets, peeping at others’ answers, and delivering short messages. These behaviours are fleeting and are not easily identified, which can cause a heavy load on the examiners’ attention. In contrast, every webcam represents each examinee’s eye during online exam. The notarisation process conducted by this study can take photo of examinees randomly for authentic identity, and take video files to monitor the whole exam process and record them in database. If an examinee is suspected of cheating, examiner can check the video files to decide after the fact whether he or she was cheating or not during the online exam. The goal is to achieve the fairness of the notarisation process during an online exam. As concluded by Asha (2008), and Doukas and Andreatos (2006), once online testing can offer a fair exam environment like that of traditional exams, examinees will feel the exam is trustworthy and their participation acceptance will increase. The Notarisation process developed in this study not only replaces the simple verification process of checking accounts and passwords, but also gives up the complicated task of fingerprint recognition in order to achieve a fair notarisation process.

Another very significant determinant which raised examinees’ Perceived Trust of online exam was the Perceived Soundness of Online Exam Policy. From the perspective of the institution level, this study found that the more sound examinees felt about online exam policy, the more reliable they felt the exam was. This also suggests that when
information and communication technology is integrated into the scenario of an exam, more and more management issues related to online exams will emerge. For this reason, administration should establish and develop a sound online exam policy in an attempt to effectively manage the operation of online testing. An example of this is examinees having to complete answers within a valid visible scope and rotational angle of the webcam with the system immediately issuing a warning ten seconds after the system cannot detect the examinee’s face; when the number of warnings amounts to three, the system will log out automatically, terminating the exam session. In addition, the types of questions used on online exams should warrant closer scrutiny in future work. In Taiwan, traditional test items include yes/no, multiple-choice, and short-answer questions, but given that there are standard answers for yes/no and multiple-choice questions, cheating behaviours for these types are more widely seen in both online and traditional exams (King et al., 2009). This study used randomised items and randomised seating as part of their strategies of online exam policy to decrease the possibility of cheating. However, smart phones, for example, have become popular as a covert device that can deliver short messages to facilitate the cheating behaviour. The best way to prevent this from happening during online exams is to redesign different types of questions suitable for different disciplines, such as “open-book” questions, which do not have standard answers. In order to prevent examinees from plagiarizing articles from other websites, the screen can be locked until the end of exam, denying access to Internet websites, such as Google, Wikipedia, and Facebook. In summary, as long as administration is able to establish a sound online exam policy to effectively manage the exam operation in addition to a fair notarisation process, examinees will think that online exams are more reliable, thus encouraging their higher participation acceptance for future online exams.

According to discussion and conclusion, this study offers several suggestions from technical and institutional levels for the development of online exams in the future.

- School administration should play the role of manager of making policy, of continuously developing contemporary online exam policy, strategy, rules, rewards and penalties, and of establishing a complete process to effectively manage the operations of the online exam.
- School administration, such as that relating to the computer centre, should play the role of monitor carrying out policies, and continue to develop more advanced equipment of information technology, which are applied in the online exam environment in order to ensure the fairness and justice of the exam process. Although the notarisation process proposed in this study can achieve fairness, information technology continues to evolve and users’ behaviours keep changing, such as with the use of smart phones combined with Apps to deliver instant messages. Therefore, only by continuously developing more advanced hardware and software of information technology and applying them into online exams can students be prevented from adopting next generation information technologies in cheating.
- Educators should reconsider the impact of IT on the pattern of traditional exam. They should thoroughly redesign all new exam types and processes based on the characteristics of online exams. For example, an open-book type of online exam allows examinees to surf the Internet, but should use a checking system to ascertain whether answers are copied; examinees should also be prohibited from using handheld mobile devices to communicate via text or spoken secret messages.

Acknowledgements

This work was supported by the National Science Council of Taiwan under contract numbers NSC-100-2511-S-161-001. The authors would like to thank all the people who took part in this study.

References


# Appendix A

## Question items used in this research

<table>
<thead>
<tr>
<th>Item statement</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived Fairness of Notarisation Process</strong></td>
<td>Self-developed</td>
</tr>
<tr>
<td>1. I feel that the notarisation process can prevent cheating behaviours in online exams.</td>
<td></td>
</tr>
<tr>
<td>2. I feel that face recording is an effective method for authenticating examinee identities in online exams.</td>
<td></td>
</tr>
<tr>
<td>3. I think that the chances of impersonation are very low if a webcam is used to photograph and record examinees.</td>
<td></td>
</tr>
<tr>
<td>4. From technical level, I think that the online exam is fair to identify examinees’ real identities.</td>
<td></td>
</tr>
<tr>
<td><strong>Perceived Soundness of Online Exam Policy</strong></td>
<td>Self-developed</td>
</tr>
<tr>
<td>1. I feel that randomizing items is a good strategy that can prevent cheating behaviours in online exams.</td>
<td></td>
</tr>
<tr>
<td>2. I feel that school provides complete online exam regulations and processes.</td>
<td></td>
</tr>
<tr>
<td>3. I would not dare to cheat as any violation of the exam rules might result in punishment from the school.</td>
<td></td>
</tr>
<tr>
<td>4. From institutional level, I think that the online exam policy is sound to manage online exams.</td>
<td></td>
</tr>
<tr>
<td><strong>Perceived Trust</strong></td>
<td>adapted from Chen, Hsieh, &amp; Tsai (2008); Ong, Lai, &amp; Wang (2004)</td>
</tr>
<tr>
<td>1. I feel that the notarisation process used to authenticate examinee identities is reliable.</td>
<td></td>
</tr>
<tr>
<td>2. I feel that the strategies used to prevent cheating behaviours in online exams are trustworthy.</td>
<td></td>
</tr>
<tr>
<td>3. I feel that the information security of online exam is creditable.</td>
<td></td>
</tr>
<tr>
<td>4. I feel that the efficacy to manage online exams is believable.</td>
<td></td>
</tr>
<tr>
<td>5. Overall, I think that the online exam is trustworthy to me.</td>
<td></td>
</tr>
<tr>
<td><strong>Participation Acceptance</strong></td>
<td>adapted from Chen, Hsieh, &amp; Tsai (2008); Davis &amp; Venkatesh (1996)</td>
</tr>
<tr>
<td>1. I am satisfied with the online exam experience.</td>
<td></td>
</tr>
<tr>
<td>2. I can accept that evaluations of my learning outcomes through the online exams.</td>
<td></td>
</tr>
<tr>
<td>3. I can accept the inconveniences that occasionally occur due to lack of familiarity with computer operations during online exams.</td>
<td></td>
</tr>
<tr>
<td>4. I can accept the eventuality that more courses will be conducting their exams online in the future.</td>
<td></td>
</tr>
<tr>
<td>5. Overall, I am accepting of participation in next online exam.</td>
<td></td>
</tr>
</tbody>
</table>
Apply an Augmented Reality in a Mobile Guidance to Increase Sense of Place for Heritage Places

Yu-Lien Chang, Huei-Tse Hou, Chao-Yang Pan, Yao-Ting Sung and Kuo-En Chang*

Department of Tourism & Leisure, Lunghwa University of Science and Technology, Taoyuan, Taiwan # 300, Section 1, Wanshou Road, Gui Shan, Taoyuan, Taiwan // Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, #43 Keelung Road, Section 4, Taipei, Taiwan // Department of Geography National Taiwan Normal University, #162, Heping East Road Section 1, Taipei, 106, Taipei, Taiwan // Department of Educational Psychology and Counseling National Taiwan Normal University, #162, Heping East Road Section 1, Taipei, 106, Taipei, Taiwan // Graduate Institute of Information and Computer Education, National Taiwan Normal University, #162, Heping East Road Section 1, Taipei, 106, Taipei, Taiwan // idachang@mail.lhu.edu.tw // hhou@mail.ntust.edu.tw // t24012@cc.ntnu.edu.tw // sungtc@ntnu.edu.tw

*Corresponding author

(Submitted December 5, 2013; Revised May 15, 2014; Accepted September 9, 2014)

ABSTRACT

Based on the sense of place theory and the design principles of guidance and interpretation, this study developed an augmented reality mobile guidance system that used a historical geo-context-embedded visiting strategy. This tool for heritage guidance and educational activities enhanced visitor sense of place. This study consisted of 3 visitor groups (i.e., AR-guidance, audio-guidance, and no-guidance) composed of 87 university students. A quasi-experimental design was adopted to evaluate whether augmented reality guidance more effectively promoted sense of place and learning performance than the other groups. The results indicated that visitors who used AR guidance showed significant learning and sense of place effects. Interviews were also employed to determine the possible factors that contribute to the formation of sense of place. Finally, a majority of the visitors who participated in the study demonstrated positive attitudes toward the use of the AR-guidance system.

Keywords

Sense of place, Augmented reality, Mobile guidance, Informal learning

Introduction

Application of sense of place and augmented reality in informal learning

Sense of place (SOP) is the combination of feelings of attachment, dependence, concern, identity, and belonging that people develop regarding a place. Since 1970, SOP has been applied in the fields of geography (Tuan, 1974; Relph, 1976; Kaltenborn, 1998; Galliano & Leoffler, 1999). SOP is also applicable in environmental psychology (Altman & Low, 1992; Jorgensen & Stedman, 2001; Nielsen-Pincus, Hall, & Force, 2010), tourism, and leisure (Williams, Patterson, Roggenbuck, & Watson, 1992; Kyle & Mowen, 2005). The various applications of SOP indicate that SOP helps establish how visitors personally experience the significance of a location (Kaltenborn, 1998). For informal learning and education (i.e., field learning and visiting heritage sites), increased SOP enables learners to develop attachment and interest toward places, which promotes learning motivation (Altman & Low, 1992; Lewicka, 2005). Furthermore, numerous scholars (Jorgensen & Stedman, 2001; Kyle & Mowen, 2005; Nielsen-Pincus et al., 2010) have indicated that the SOP concept encompassed visitors’ affections toward and cognitions about a place. Based on previous studies defining SOP (Jorgensen & Stedman, 2001; Nielsen-Pincus et al., 2010), the following three constructs for SOP were incorporated in this study: (1) place attachment (PA), (2) place dependence (PD), and (3) place identity (PI), which represent the affective, behavioral, and cognitive relationships, respectively, that people develop with places.

In studies on heritage tourism, Prentice, Guerin, and McGugan (1998) indicated that visitors’ affective and cognitive experiences regarding heritage sites were associated with the quality of the interpretative media provided for those places. Specifically, interpretations of heritage sites promote the development of PI (Uzzell, 1996), and interpretive activities at recreation sites enhance SOP formation and encourage the conservation of local cultural heritage (Stewart, Hayward, & Devlin, 1998). Furthermore, Knudson, Cable, and Beck (2003) suggested that the aim of interpretation is to assist visitors in developing SOP. Overall, high-quality historical and geographical guidance and
Researchers have often stressed enhancing the functions of museums, such as designing educational projects connected to both learners’ formal classroom education and informal out-of-school learning. Researchers have often stressed enhancing the functions of museums, such as designing educational projects connected to both learners’ formal classroom education and informal out-of-school learning. (Cox-Petersen, Marsh, Kisiel, & Melber, 2003; Dierking, Falk, Rennie, Anderson, & Ellenbogen, 2003). Technology is currently used extensively in heritage site and museum guidance. Specifically, augmented reality (AR) is one of the primary technological applications used to assist historical site or museum guidance activities. AR is the combination of virtual and real objects in a real environment (Azuma, Baillot, Behringer, Feiner, Julier, & MacIntyre, 2001; Chang, Chang, Hou, Sung, Chao, & Lee, 2014; Zhang, Sung, Hou, & Chang, 2014). AR is the technology that overlays digital information on objects or places in a real world for the purpose of enhancing the user experience (Berryman, 2012; Chang et al., 2014; Zhang et al., 2014). Studies have shown that visitors perceive AR guidance activities as interesting, fun, challenging, and a method of achieving first-hand experience. Consequently, AR guidance activities may increase people’s motivation to visit and learn knowledge (Dunleavy, Dede, & Mitchell, 2009; Klopfer & Squire, 2008; Mulloni, Wagner, & Schmalstieg, 2008; Liu, Tan, & Chu, 2009; Priestnall, 2009; Luckin & Stanton Fraser, 2011; Chang et al., 2014; Zhang et al., 2014).

AR enhances location-based learning because AR helps learners connect historical sites with their prior knowledge and provides additional new knowledge relevant to the sites. For example, AR enables visitors to observe and experience comparisons between cultural artifacts that have disappeared and existing artifacts in a site. This may enhance visitors’ understanding of the history of cultural heritage and increases the level of immersion (Portalés, Lerma, & Pérez, 2009; McCall, Wetzel, Löschner & Braun, 2011). AR technology includes information on people and buildings on the site and allows users to switch between historical periods and observe the different appearances of a city throughout time. This combination of additional information and actual scenes enhances people’s senses of reality and presence in a certain place.

However, the application of AR in guidance activities demonstrates areas for improvement (Billinghurst et al., 2003; Dunleavy et al., 2009; Wang & Chen, 2009; McCall et al., 2011). To ensure that visitors view the additional information presented through AR in a specific visual field, the technology is designed to attract visitor attention. Hence, visitors tend to focus excessively on the content of the AR system and the additional information in the AR, thereby neglecting the physical surroundings and environment (Billinghurst et al., 2003; Dunleavy et al., 2009; Wang & Chen, 2009). McCall et al. (2011) contended that for some visitors, the sense of presence shaped by the AR environment decreases and cannot be preserved after a guidance activity is completed and vanishes as visitors proceed to their next destination.

Human–computer–context interaction and historical geo-context-embedded visiting

Sung, Chang, Hou and Chen (2010a) indicated that the planning and application of a mobile guide should accommodate the overall learning context of the guided environment, taking into consideration the visitor, the visitor’s companions, historical exhibits, and the cultural or social implications of the exhibits. The interface and content of a mobile guide plays a significant role in effectively increasing participative behaviors (e.g., visitors pay more attention to exhibits and spend more time at them). The design of the mobile guide should fully support the interaction between the visitors and these aforementioned aspects, thus forming a “human–computer–context” interaction (HCCI) (Sung et al., 2010a). This study proposes an HCCI framework to design a mobile guide for enhancing interactions and stimulate motivation. The HCCI framework has two features. The first emphasizes that the design and application of the mobile guide should encompass the situation within a learning environment, including the visitors, the exhibits, and the cultural or social meanings behind the exhibits (Falk & Dierking, 2000). Secondly, a mobile guide, as a tool with embedded information, is intended to facilitate a visitor’s interaction with all aspects of the context about the exhibits. There are two types of interaction. The traditional type is the visitor–computer interaction, in which visitors may retrieve or save information in the mobile device. However, the potential for the mobile device to connect the exhibits and learners is minimized if it only provides a visitor–computer interaction. The other type is the human–computer–context interaction. In this type of interaction the visitors not only pay attention to the physical features of the exhibits, but also interact with the historical background or cultural context, which provides a more in-depth visiting experience (Bain & Ellenbogen, 2002). Therefore, this
type of mobile guide not only support the interaction between visitors and exhibits, but it also reinforces the interaction between the visitors and the context behind the exhibits (e.g., being aware of historic and social perspectives when viewing the heritage site).

Situated cognition is the theoretical foundation for mobile guidance systems based on the HCCI framework. Situated cognition assumes that knowledge and learning are produced through a cognitive learning process as visitors interact within a given context (i.e., the surrounding social or physical environmental context) (Brown, Collins, & Duguid, 1989; Greeno, Collins, & Resnick, 1996). The mobile learning guidance with HCCI can enrich visitor experience by increasing their interactions when interpreting cultural or historical contexts and situational or artistic conceptions of exhibit (Beck & Cable, 2002). In light of this, the limitation underlying the planning for AR mobile guide activities lies in the inability to balance a visitor’s attention distribution between the additional information provided by AR and the physical scenes, causing them to focus their attention on the “human–computer (guide system)” excessively, and to ignore the more important human-context (exhibition and local context)” in the real environment. In other words, it cannot induce the HCCI. The design of an AR mobile guide for exhibitions should therefore emphasize the coupling the information provided through AR and the actual environment (Klopf & Squire, 2008; Cheng & Tsai, 2013); that is, enhance the interaction between the additional information and the real physical scenes.

To develop a guidance system that incorporates HCCI for heritage learning, this study adopted a historical geo-context-embedded visiting (HGCEV) guidance strategy. The first step of HGCEV is to find out the past geographical and historical information about the heritage site, and then to establish its geographical and historical context. When visitors visit the heritage site, the context allows them to feel interested in and interact with the heritage site, and further to establish the interaction among visitors, the heritage site itself, and the geographical and historical context of the heritage site. This is the purpose of using the HGCEV strategy to achieve HCCI. The HGCEV strategy may involve the integration of affective experience, conative behavior, and cognitive knowledge of history and geography with guidance regarding heritage sites. This is accomplished by constructing the historical narratives and local implications of the heritage sites into the AR system. This accentuates the individuality of different places, immerses and integrates visitors in the local context, increases visitor interest, improves visitor familiarity with, sense of belonging to, and identity with the geographical location, and facilitates the active exploration of the significance and value of heritage places.

These features compensate for the disadvantages of AR guidance application by establishing HCCI (Dunleavy et al., 2009). Few studies have examined the use of AR as a guidance tool to increase SOP and learning performance. This study integrates the HGCEV strategy into a mobile AR guidance system to construct a HCCI guidance model and applies this model to heritage guidance and interpretive activities. In addition to guidance system development, the experimental objectives of this study are as follows:

- Examine the differences in the learning achievement effects for visitors who used the AR mobile-guidance device, audio guidance, or no assistive guidance during a heritage visit.
- Examine and analyze the different effects of the three guidance modes on the three SOP constructs (PA, PD, and PI) and overall SOP during a heritage visit.
- Use interviews to understand the attitudes of visitors toward the three guidance modes and the possible contributing factors in the formation of the three constructs of SOP (PA, PD, and PI).

**System design**

The AR guidance system on heritage places was developed by integrating HGCEV strategy which promotes the PA, PD, and PI constructs of SOP. In investigating these three constructs, qualitative study with contextual analysis and in-depth interviews was conducted. The purposive sampling method was applied to conduct interviews with 21 residents (9 women and 12 men) and 18 visitors (9 women and 9 men) to heritage sites in Tamsui District in Taiwan to identify possible factors of SOP constructs. The factors about PA, PD, and PI are listed in Table 1. For example, for the PA construct, the factors include place affection and place interest.

To effectively construct the additional information displayed by AR during a guidance activity, it is presented according to the interpretation principles results shown in Table 1 proposed by Tilden (1977) and Beck and Cable (2002). For example, the presentation of additional information to promote PA factors was based on two interpretation principles: (1) ensuring that visitors associate personal characteristics and experiences with the heritage
on site and creating an affective bond between visitors and heritage site; and (2) increasing visitors’ interest and allowing them to experience the joy of understanding a place. The forms of media used to present the supplemental information are also described in Table 1. For example, in place affection, films, images, and songs related to past living experiences were used to promote visitor attachment to the place.

Table 1. Theoretical framework for the AR-guidance system

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Nature of content</th>
<th>Formation factors</th>
<th>Interpretive principles</th>
<th>The presentation media and styles</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA (Place Attachment)</td>
<td>Affective</td>
<td>Place affection</td>
<td>Ensuring that visitors associate personal characteristics and experiences with heritage at the site and creating an affective bond between visitors and the heritage site</td>
<td>Movie images and songs related to past living experiences in a place are used.</td>
</tr>
<tr>
<td>Place interest</td>
<td></td>
<td>Increasing visitor interest and allowing visitors to experience the joy of understanding a place</td>
<td>Images, audios, and texts that interact with the physical environment of a place are provided.</td>
<td></td>
</tr>
<tr>
<td>SOP (Sense of Place)</td>
<td>Conative Behavior</td>
<td>Place uniqueness</td>
<td>Demonstrating the uniqueness of a place</td>
<td></td>
</tr>
<tr>
<td>PD (Place Dependence)</td>
<td></td>
<td>Encourages active exploration</td>
<td>Encouraging visitors to actively explore and create meaning and significance</td>
<td></td>
</tr>
<tr>
<td>PI (Place Identity)</td>
<td>Cognitive</td>
<td>Historical and geographical implications</td>
<td>Representing historical and geographical information and revealing its fundamental significance to enhance the connection between historical and geographical evolution and on-site heritage</td>
<td>Texts, audios, and images illustrating the geographical and environmental shifts and historical context of heritage sites are presented.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inspires PI</td>
<td>Encouraging visitors to perceive their self-worth and understand the relationship of on-site heritage to a place or setting and inspiring PI</td>
<td>Texts, audios, and images are combined with the heritage site to depict its geographical and environmental shifts and historical background.</td>
</tr>
</tbody>
</table>

In this study, the mobile carrier hardware used for the AR mobile guidance system was the ASUS Eee Pad Transformer TF101 (Android 3.2). The streamlined viewport strategy algorithm developed by Zhang, Hou, and Chang (2012) was employed as the identification method of the AR guidance system. It has the capacity to load additional information or interpretations regarding historical or geographical contexts when visitors aim the camera of a tablet computer at a historical site. This enables convenient operation for the visitors. The proposed system includes the following functions:

- **Historical site guidance:** This function enables visitors to identify the historical site or heritage site using the device’s camera and to browse or listen to guidance content that corresponds to the historical site. The Figure 1a depicts a visitor using the AR system to identify an image of the historical structure, and the Figure 1b shows the visitor using the AR system to view and listen to the information provided by the system after the structure has been identified. Figure 2 shows how students read the information of AR system about a historical site. The location selected for this study was the Tamsui District, which is a location in Taiwan that contains numerous historical sites and heritage sites. The system contains 14 guidance and interpretation locations. The guidance and interpretation principles and presentation style were developed according to the guidelines presented in Table 1, and the contents were reviewed and assessed by two experts on the Tamsui historical sites, both historical site museum curators.
Figure 1. A visitor using the AR system to identify a historical site (a) and viewing the information provided by the AR system (Former British consular residence, b)

Figure 2. Students read the information on the AR system about the historical site

Figure 3. The zoom in and zoom out functions for observing specific areas of the historical site
• Historical site observation: This function guides visitors in identifying specific parts or areas of the historical site. The “Zoom in” and “Zoom out” functions of the device interface enable the observation of specific areas of the historical site and provide detailed interpretive information on the areas. Figure 3 presents a schematic of how the system guides visitors to employ the zoom-in and zoom-out functions to view specific areas of the historical site.

**Methodology**

This study examined the effects that visitor use of AR guidance systems had on overall SOP effects; PA, PD, and PI effects; and learning achievement effects before and after experimental processing. This study also investigated the potential factors that contribute to the formation of SOP in visitors and their attitudes toward using guidance systems.

**Participants**

The participants in this study were 87 first-year university students of the Department of Tourism and Leisure from three classes in Taiwan. Among these, 31 were male and 56 were female subjects, and the average age was 19 years. Participants were grouped via random sampling. Students in experimental group 1 used the AR mobile guidance system (the AR-guidance group; 31 students). Students in experimental group 2 used audio guidance (the audio-guidance group; 32 students). Students in the control group did not use any guidance (no-guidance group; 24 students). The difference between the AR-guidance and audio-guidance groups was that the audio-guidance group was provided with audio guidance only, although the content was identical to that provided to the AR-guidance group. Conversely, the AR-guidance group received both visual and audio guidance simultaneously.

**Experimental design**

The study employed a quasi-experimental design. The independent variable was the guidance mode group (AR-guidance group, audio-guidance group, and no-guidance group). The dependent variables were learning achievement, the three SOP constructs (PA, PD, and PI), and overall SOP effect. This study also employed interviews to understand visitor attitudes toward the guidance system and the factors that contribute to SOP formation. Participants in experimental groups 1 and 2 used the AR-guidance system and the audio-guidance system, respectively. They operated the devices independently for 90 minutes by following the system instructions. Participants without the assisting guidance devices visited the historical sites for 90 minutes. Prior to the experiment, all participants completed a SOP scale survey and a learning achievement pretest regarding Tamsui cultural and historical sites (the duration for each test was 15 minutes). Participants in the experimental groups received additional training in operating the guidance system and were able to adjust the device appropriately (approximately 15 minutes). This procedure ensured that the effects of technology barrier factors were removed. Subsequently, the three groups proceeded with their independent visiting activities (90 minutes for each group). After the tour, all groups were required to take the learning achievement posttest and complete the SOP scale. Twenty visitors from each group were randomly sampled for a follow-up interview.

**Instruments**

*Learning achievement test*

The learning achievement test was a paper-and-pencil test that was conducted both as a pretest and a posttest. The purpose was to examine whether participants had increased factual knowledge of the historical sites and heritage of Tamsui and to determine learning achievement effects after employing various guidance modes. The test contained 15 multiple-choice questions. A Tamsui historical site and heritage expert and a university teacher currently teaching courses related to guidance in travel culture provided opinions and revisions regarding the proposed test questions. The contents of the pretest and posttest were identical. The same test was used for the pilot test, which was
conducted with 50 first-year university students. The coefficient of the reliability analysis was 0.503 with an internal consistency.

**SOP scale survey**

This study used the 5-point (1 to 5) SOP scale proposed by Jorgensen and Stedman (2001) to measure each SOP construct and the overall SOP effects in all participants following historical site and heritage guidance. These items were modified based on previous research on the measurement items used for the three place constructs (Jorgensen & Stedman, 2001; Kyle & Mowen, 2005; Nielsen-Pincus et al., 2010). This scale was piloted with 50 first-year university students, and the reliability coefficients of SOP, PA, PD, and PI were 0.85, 0.712, 0.687, and 0.731, respectively.

**Interview questions**

Interview opinions and feedback were employed to understand the potential factors that contribute to the formation of SOP and the system participant attitudes of the various groups after the guidance activity. The portion of the interview regarding SOP formation factors was designed to understand the affective experience, conative behavior, and cognitive knowledge that the participants developed during the visiting activity. In addition, the interview questioned participants regarding the influence that encountering certain scenarios or conditions had during the activity process, and participant attitudes toward using the system were also examined. The interview questions included the degree to which the visitors enjoyed the activity or using the system, whether the activity or system increased their level of interest, and whether participants increased their knowledge. In addition, questions concerning learning motivation and opinions regarding the visit were included.

**Results**

**Learning achievement effects analysis**

A one-way analysis of covariance (one-way ANCOVA) was performed to compare the posttest scores of the experimental groups and the control group after the influence of the pretest scores was removed. The results are presented in Table 2. After the effect of the pretest scores on the posttest scores was removed, the between-group scores were significantly different ($F = 10.497$, $p < 0.05$). The results of the ANCOVA indicated that after the experimental processing, the different guidance modes had a significant effect on the learning achievements of the participants, and the learning achievements among the three groups differed significantly. A pairwise least significant difference (LSD) test was conducted. The results demonstrated that the posttest scores of the no-guidance and audio-guidance group were not significantly different ($F = 10.497$, $p > 0.05$) but that the posttest score of the AR-guidance group was significantly higher than those of the no-guidance group ($F = 10.497$, $p < 0.05$) and the audio-guidance group ($F = 10.497$, $p < 0.05$). These results indicate that the participants in the AR-guidance group demonstrated superior learning achievements.

<table>
<thead>
<tr>
<th>Variables</th>
<th>AR-guidance group ($n=31$)</th>
<th>Audio-guidance group ($n=32$)</th>
<th>No-guidance group ($n=24$)</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning effects</td>
<td>Pretest: 39.41 Posttest: 65.84</td>
<td>Pretest: 45.47 Posttest: 56.28</td>
<td>Pretest: 43.33 Posttest: 55.31</td>
<td>10.497*</td>
</tr>
</tbody>
</table>

* $p < 0.05$.
Analysis of the SOP effect

The SOP scale results for the three groups were used to analyze the effects for the three SOP constructs (PA, PD, and PI) and the overall SOP. A one-way ANCOVA was conducted to examine the differences in the SOP-scale posttest scores of the three groups after the effect of the SOP pretest scores was removed. The results of the ANCOVA are presented in Table 3. The results show that guidance mode did not have a significant influence on PA, and no significant differences between the groups were found ($F = 1.622, p > 0.05$). A pairwise LSD test was performed on the PD scores, but no significant difference was found between the posttest scores of the no-guidance group and the audio-guidance group ($F = 3.410, p > 0.05$). However, the posttest scores of the AR-guidance group were significantly higher than those of the audio-guidance and no-guidance groups ($F = 3.410, p < 0.05; F = 3.410, p < 0.05$). These results indicate that the AR-guidance group experienced greater PD effects. In addition, a pairwise LSD test was performed on the PI scores, but no significant difference was found between the posttest scores of the no-guidance and audio-guidance groups ($F = 4.793, p > 0.05$). Nevertheless, the posttest scores of the AR-guidance group were significantly higher than those of the audio-guidance and no-guidance groups ($F = 4.793, p < 0.05; F = 4.793, p < 0.01$). These results indicate that the AR-guidance group experienced a greater PI effect. Finally, a pairwise LSD test was performed on the overall SOP scores, but no significant differences were found between the posttest scores of the audio-guidance and no-guidance groups ($F = 4.306, p > 0.05$). The posttest scores of the AR-guidance group were significantly higher than those of the audio-guidance and no-guidance groups ($F = 4.306, p < 0.05; F = 4.306, p < 0.05$). These results indicate that the AR-guidance group experienced a greater overall SOP effect.

<table>
<thead>
<tr>
<th>Variables</th>
<th>AR-guidance group (n = 31)</th>
<th>Audio-guidance Group (n = 32)</th>
<th>No-guidance group (n = 24)</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
<td>Posttest</td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td>3.1720</td>
<td>3.7742*</td>
<td>3.0521</td>
<td>3.3542</td>
</tr>
<tr>
<td>PI</td>
<td>3.1097</td>
<td>3.6839</td>
<td>3.0813</td>
<td>3.3250</td>
</tr>
</tbody>
</table>

*p < 0.05.

Analysis of the interview results

According to Table 1, the formation factors of SOP are elements of its three constructs, PA, PD, and PI, and other nature content. The interviews examined the whether different factors formed SOP, as shown in Table 4. The interview results were categorized into several contributing factors and then computed (Table 4). For example, 20 visitors from the AR-guidance group were interviewed, 18 indicated that place uniqueness contributed to their SOP formation. Therefore, place uniqueness has a 90% possibility of contributing to SOP formation. The results shown in Table 4 suggest that, compared to the audio-guidance and no-guidance group, more participants in the AR-guidance group indicated “place uniqueness,” “historical and geographical implications,” and “inspires place identity” as possible contributing factors for SOP.

To understand the attitudes of visitors in different groups, the interviews were organized and subsequently analyzed using one-way analysis of variance to obtain independent samples. The results indicated significant between-group differences regarding attitudes, including the level of enjoyment, interest, knowledge acquired, and learning motivation ($F = 125.659, p = 0.000 < 0.05; F = 98.335, p = 0.000 < 0.05; F = 129.408, p = 0.000 < 0.05; F = 91.791, p = 0.000 < 0.05$; Table 5). Multiple comparison analysis and post hoc comparisons with the LSD test were performed to examine agreement levels for the four attitudes among visitors from each group. The results showed that visitors in the AR-guidance group exhibited greater levels of agreement compared with that of the visitors in the audio-guidance and no-guidance groups (Table 6).
Table 4. Potential factors that contribute to SOP formation

<table>
<thead>
<tr>
<th>SOP constructs</th>
<th>PA</th>
<th>PD</th>
<th>Historical and geographical implications</th>
<th>Inspires PI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Place affection</td>
<td>Place interest</td>
<td>Place uniqueness</td>
<td>Encourages active exploration</td>
</tr>
<tr>
<td>AR-guidance group (n = 20) Response (%)</td>
<td>20 (100%)</td>
<td>20 (100%)</td>
<td>18 (90%)</td>
<td>19 (95%)</td>
</tr>
<tr>
<td>Audio-guidance group (n = 20) Response (%)</td>
<td>20 (100%)</td>
<td>20 (100%)</td>
<td>6 (30%)</td>
<td>18 (90%)</td>
</tr>
<tr>
<td>No-guidance group (n = 20) Response (%)</td>
<td>20 (100%)</td>
<td>20 (100%)</td>
<td>2 (10%)</td>
<td>18 (90%)</td>
</tr>
<tr>
<td>Total %</td>
<td>60 (100%)</td>
<td>60 (100%)</td>
<td>26 (43%)</td>
<td>55 (92%)</td>
</tr>
</tbody>
</table>

Table 5. ANOVA of the participant attitudes

<table>
<thead>
<tr>
<th>Variables</th>
<th>AR-guided group (n = 20)</th>
<th>Audio-guided group (n = 20)</th>
<th>No-guided group (n = 20)</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyment</td>
<td>4.9</td>
<td>3.5</td>
<td>3.0</td>
<td>125.659*</td>
</tr>
<tr>
<td>Interest</td>
<td>4.8</td>
<td>2.5</td>
<td>3.2</td>
<td>98.335*</td>
</tr>
<tr>
<td>knowledge acquired</td>
<td>4.95</td>
<td>2.75</td>
<td>3.2</td>
<td>129.408*</td>
</tr>
<tr>
<td>learning motivation</td>
<td>4.95</td>
<td>3.45</td>
<td>2.8</td>
<td>91.791*</td>
</tr>
</tbody>
</table>

*p < 0.01.

Table 6. Post-hoc comparisons of the participant attitudes

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>(I) group</th>
<th>(J) group</th>
<th>Average difference (I–J)</th>
<th>Standard error</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyment</td>
<td>AR-guided group (n = 20)</td>
<td>Audio-guided group (n = 20)</td>
<td>1.4*</td>
<td>0.124</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Audio-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>1.9**</td>
<td>0.124</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>Audio-guided group (n = 20)</td>
<td>-1.4*</td>
<td>0.124</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>0.5**</td>
<td>0.124</td>
<td>0.000</td>
</tr>
<tr>
<td>Interest</td>
<td>AR-guided group (n = 20)</td>
<td>Audio-guided group (n = 20)</td>
<td>2.3**</td>
<td>0.167</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>1.55**</td>
<td>0.167</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Audio-guided group (n = 20)</td>
<td>AR-guided group (n = 20)</td>
<td>-2.3**</td>
<td>0.167</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Audio-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>-0.75**</td>
<td>0.167</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>Audio-guided group (n = 20)</td>
<td>-1.55**</td>
<td>0.167</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>0.75**</td>
<td>0.167</td>
<td>0.000</td>
</tr>
<tr>
<td>knowledge acquired</td>
<td>AR-guided group (n = 20)</td>
<td>Audio-guided group (n = 20)</td>
<td>2.20**</td>
<td>0.167</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>1.75**</td>
<td>0.144</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Audio-guided group (n = 20)</td>
<td>AR-guided group (n = 20)</td>
<td>-2.2**</td>
<td>0.144</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Audio-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>-0.45**</td>
<td>0.144</td>
<td>0.003</td>
</tr>
<tr>
<td>learning motivation</td>
<td>AR-guided group (n = 20)</td>
<td>Audio-guided group (n = 20)</td>
<td>-1.5**</td>
<td>0.162</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>1.5**</td>
<td>0.162</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Audio-guided group (n = 20)</td>
<td>AR-guided group (n = 20)</td>
<td>2.15**</td>
<td>0.162</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Audio-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>0.65**</td>
<td>0.162</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>Audio-guided group (n = 20)</td>
<td>-2.15**</td>
<td>0.162</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>AR-guided group (n = 20)</td>
<td>No-guided group (n = 20)</td>
<td>-0.65**</td>
<td>0.162</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*p < 0.01. **p < 0.001.
Discussion and conclusions

In the study, the experimental processing of the three groups of visitors using different guidance modes confirmed that different guidance modes produced varying learning achievements. Among the groups, participants in the AR-guidance group appeared to learn more than other groups. Moreover, the analysis of the interviews showed that learning results increased because the majority of the participants in the AR-guidance group indicated that the guidance system helped them acquire greater knowledge of the historical sites and increased their learning motivation. The AR-guidance system provided audio interpretations and integrated supplemental information, including text and images pertaining to the history and geography of the sites. This mode effectively increases visitors’ knowledge of historical sites, enhances their motivation to learn, and facilitates interaction among the visitors, the guidance system, and historical sites (Klopfer & Squire, 2008). It also helps visitors construct a memory of the place. These results confirm the necessity of human–computer–context interaction (Sung et al., 2010a; Sung, Hou, Liu, & Chang, 2010b). Numerous studies have indicated that AR-assisted teaching tools promote learning performance (Dunleavy et al., 2009; Liu et al., 2009; Chang et al., 2014; Zhang et al., 2014). AR encourages users to participate and constructs a learning environment that integrates supplemental information and reality. In addition, this learning environment is realistic and innovative and is integrated with supplemental information (Dunleavy et al., 2009; Klopfer & Squire, 2008; Mulloni et al., 2008; Luckin & Stanton Fraser, 2011). Consistent with the results of previous studies, the system developed in the study showed significant facilitation effects on the overall learning achievements of the participants.

Regarding the SOP effects, the results of this study indicate that, aside from the PA construct, participants in the AR-guidance group scored higher on the PD and PI constructs and the overall SOP than the participants of the audio-guidance and no-guidance groups. Furthermore, the posttest scores between the audio-guidance group and no-guidance group were not significantly different. Nevertheless, the PA effect did not differ between the three groups, which seems inconsistent with the findings of Altman & Low (1992) and Lewicka (2005). In those studies, PA was enhanced with increases in knowledge about a city’s past. However, the PA conceptualized in these two studies was a compound concept that incorporated the PA and PI constructs described in the study. The findings of the study concur with the results of Altman & Low (1992) and Lewicka (2005) when the increased PI construct effect determined in this study is considered. This study has determined that the students of AR-guidance group using AR system designed based on the HGCEV strategy experienced a greater PD and PI. Students are likely to focus and acquire additional knowledge about a place when they perceive the site as being unique or when they develop increased site attachment and identification. This promotes learning motivation and positively influences learning achievements. The analysis of learning effects performed in this study revealed that the learning achievement effects were superior in students in the AR-guidance group than in students in the other groups.

Regarding factors that contribute to SOP formation, 100% of the participants in the AR-guidance group responded positively regarding place affection and place interest. Most (90%) of the visitors in the AR-guidance group stated that Tamsui was unique and irreplaceable. Moreover, 95% of visitors were willing to engage in active exploration. The AR-guidance group expressed more significantly positive responses than the audio-guidance and no-guidance groups.

In summary, this study found that AR enhanced learner SOP and learning achievements (including learning motivation and learning achievements in knowledge of local history), and effectively improved learning achievements when applied in learning related to heritage sites. This tool can therefore be applied to enhance learning achievements. These findings are consistent with those of numerous previous studies that applied AR to improve learning achievements (Ibáñez et al., 2014; Zhang et al., 2014; Dunleavy et al., 2009; Liu et al., 2009; Martín-Gutiérrez et al., 2010).

Other issues that warrant consideration include ergonomic or human-factor concerns, including the sensitivity of AR to light, camera angles of the portable, handheld devices. These challenges must be overcome in the future development of mobile AR application devices. Furthermore, this study identified a crucial effect of AR-guidance during visits: When focusing on historical site interpretations during outdoor activities, visitors were susceptible to distractions from external objects and engaged in fewer discussions and interactions about the historical sites with their companions. Consequently, the system tested in this study limited interpersonal interactions during visits to historical and heritage sites. Therefore, incorporating a systemic, planned question and answer strategy and other learning strategies into future AR guidance and integrating appropriate cooperative learning, situated tasks (Bruner,
1985; Sánchez & Olivares, 2011) is suggested to encourage interactions among visitors and ensure peer–computer–context interaction (Sung et al., 2010a; Sung et al., 2010b).

Future studies may continue to apply and extend the HGCEV strategy in AR-guidance systems to improve learning and SOP effects. Additional interactive strategies (e.g., games and role-playing) can be incorporated to enhance the first-hand experiences and interactions of users.

Acknowledgments

This study was supported by grants from the National Science Council, Taiwan, Republic of China (Contract Numbers: NSC-102-2511-S-011-001-MY3, NSC-100-2628-S-011-001-MY4, NSC-99-2511-S-003-026-MY3 and NSC-99-2511-S-011-007-MY3) and “The International Research-Intensive Center of Excellence program” (NSC103-2911-I-003-301) of the National Taiwan Normal University and the National Science Council, Taiwan, R.O.C. We also appreciate Mr. Jia Zhang and Ms. Chia-Tzu Chang for their assistance of related experimental procedures.

References


Conceptual Tutoring Software for Promoting Deep Learning: A Case Study

Angela Stott* and Annemarie Hattingh
School of Education, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa // stottae@ufs.ac.za // annemarie.hattingh@uct.ac.za

*Corresponding author

(Submitted December 21, 2013; Revised July 15, 2014; Accepted July 25, 2014)

ABSTRACT

The paper presents a case study of the use of conceptual tutoring software to promote deep learning of the scientific concept of density among 50 final year pre-service student teachers in a South African university. Individually-paced electronic tutoring is potentially an effective way of meeting the students’ varied learning needs within the limited time available for content teaching in the pedagogically focused course. Both qualitative and quantitative data were collected in the forms of engagement patterns, surveys, focus-group discussions, and pre- and post-test scores. These were used to formulate a rich description in answer to the questions “Is deep learning evident when student teachers engage with a module on density using conceptual tutoring software?”, and “What design features of this module are effective in promoting deep learning?”. The findings suggest that conceptual tutoring software can be effective in promoting deep learning among student teachers. The sub-step design, branched structure, and conceptual approaches used in the tutoring design, and the inclusion of a written task and a test, were found to be effective in promoting deep learning.

Keywords

Electronic tutor, Conceptual, Software, Deep learning, Science pre-service teacher education, PGCE, Natural sciences, Density

Introduction

The Postgraduate Certificate in Education (PGCE) Natural Sciences (NS) methods course is part of a single-year program which qualifies graduates to teach in grades 4 to 9 in South Africa as generalists of various subjects. Little of the class contact-time can be used to develop content knowledge since the course has a pedagogical focus. Also, the large range in content knowledge among the students makes it difficult to pitch content teaching at a level beneficial to all the students. Therefore, individually-paced electronic tutoring is potentially an effective way of meeting the students’ varied learning needs. There is evidence that conceptually-oriented tutoring software can be effective in helping students to engage in deep conceptual learning (e.g. She & Liao, 2010). The existing literature on use of such software to promote deep conceptual learning, including research previously done on the particular software used in this study, iQuiz (Stott & Case, 2014), has almost exclusively been quantitative and experimental in nature. In this case study we sought to address this gap in the literature through a rich, primarily qualitative study of students’ learning experiences and behavioural choices during their extended engagement with conceptual tutoring software. This occurred during a module, within the PGCE NS methods course, designed to teach the scientific concept of density. The research is guided by the questions: (1) “Is deep learning evident when student teachers engage with a module on density using conceptual tutoring software?,” (2) “What design features of this module are effective in promoting deep learning?”

Literature review

Effective tutoring provides individualised formative feedback of a small enough grain-size to support a learner to remain engaged in learning for long enough to think through all the steps required for mastery (VanLehn, 2011). Grain size refers to ‘the amount of reasoning required of participants between opportunities to interact’ (p. 203). VanLehn calls electronic tutors (e-tutors) which offer individualised feedback of a very fine grain size sub-step intelligent tutoring systems (ITSs). A number of ITSs have been developed and researched, with generally high levels of success reported. Almost all of these ITSs provide support for learners solving problems (Mitrovic & Weerasinghe, 2009), particularly algorithmic-based problems. A few guide learners to solve more ill-structured problems (Fournier-Viger, Nkambou, & Nguifo, 2010) and a few tutor conceptual understanding. These are scarce
due to the complexity associated with their creation in terms of both software programming and content which runs on the software (Mitrovic, Martin, & Suraweera, 2007).

In this study we use software which we refer to as a conceptual e-tutor, called iQuiz (iquiz.uct.ac.za), which was developed at a South African university. iQuiz can be created to provide formative feedback of a very fine grain size which differs for different learner responses, and therefore meets VanLehn (2011)'s criteria for potential high effectiveness and inclusion in the ITS category. Most authorities, however, require software to exhibit artificial intelligence (AI) to be called an ITS (e.g. Aleven, McLaren, Sewall, & Koedinger, 2009). iQuiz does not use AI features, although it can be pre-programmed to respond differently to different user choices at each slide. iQuiz consists of a number of multiple-choice-question-, and information-, slides which are connected to one another by hyperlinks. Some of these slides form a central backbone, traversed if all questions are answered correctly. Other slides form side branches, designed to respond appropriately to a user’s incorrect responses in what could be called an electronic dialogue. Content-creation is time-consuming, preferably requires advanced pedagogical and conceptual content knowledge and skill, but does not require programming knowledge and is highly flexible. Figure 1 represents part of the e-tutor. Each of the ovals refers to a slide and the connecting arrows to hyperlinks.

**Figure 1.** A representation of a part of an iQuiz e-tutor

Although previous research on iQuiz (Stott & Case, 2014) suggests that it can be an effective conceptual e-tutor, this research was done in the context of a short-duration engagement during an in-service teacher workshop. This did not provide information about how effective the software would be during more extended courses where students have greater flexibility about their engagement choices and tend to use information and communication technologies (ICT) in limited and shallow manners (Selwyn, 2007). Therefore, in this research we sought to understand how student teachers use this e-tutor as a resource in their learning for understanding.

**Deep learning**

Deep learning refers to learning for understanding (Schwartz, Lindgren, & Lewis, 2009). Mayer (2009)'s Theory of Multimedia Learning can be used as a lens for recognising deep learning. According to this theory, the small size and crucial role in learning, of working memory, is the limiting factor in learning. Three kinds of cognitive processing occur in working memory: Extraneous, essential, and generative processing (Mayer, 2009). For deep learning to
occurrence, extraneous processing must be minimised, and essential and generative processing supported. Extraneous processing is cognitive activity which is not helpful to the learning process. In the context of learning with ICT, the possibility to *game the system*: Making progress within the software while avoiding learning (Mulder, Burleson, Van de Sande, & VanLehn, 2011), is a potential source of extraneous processing. During essential processing the learner selects and organises appropriate information in order to comprehend the information presented. This can be detected by the learner’s performance on recall-type questions, i.e., those requiring responses similar to the presented information. During generative processing the learner integrates knowledge they have comprehended during essential processing with their prior knowledge, thus developing understanding. This can be detected by the learner’s performance on transfer-type questions, i.e., those requiring learners to apply their knowledge to contexts which are not similar to the presented information.

The presence of effective self-regulation in learning is also associated with engagement in deep learning (Blom & Severiens, 2008). Self-regulated learning occurs under the students’ metacognitive control and intrinsic motivation (Schraw, Olafson, & VanderVelde, 2012). Models of self-regulated learning include an iterative internal feedback loop by which a student evaluates his or her learning and checks this evaluation against goals, rates of progress, and external feedback (Dweck, 2002). This may result in changes in affect, including self-efficacy, and behavioural choices. The more effective a student is at self-regulation, the more appropriately they are able to calibrate their self-assessment against external assessment and alter affect and behaviour toward attaining their goals (Nicol & Macfarlane-Dick, 2006). The intrinsic motivation displayed by such learners may result from a mastery-oriented epistemology (Schraw et al., 2012): The belief that the purpose of obtaining knowledge is to master a domain rather than to achieve extrinsic rewards. Mastery of the domain is associated with an increase in self-efficacy (Zimmerman, Bandura, & Martinez-Pons, 1992). Effective self-regulation is particularly critical to ensuring the sustained engagement necessary for deep learning in the context of learning using ICT, since students can end a learning session at will. An e-tutor’s design features can affect the likelihood of this sustained engagement occurring or not (D’Mello, Olney, Williams, & Hays, 2012). The nature and timing of the feedback the ICT provides is particularly influential in affecting the extent to which learners engage in self-regulation, although optimal design features are still poorly understood (Narciss, 2013).

**Learning the scientific concept of density**

The scientific concept of density tends to be poorly understood by children and adults (Duckworth, 1986). Dawkins, Dickerson, McKinney, and Butler (2008) found that middle school student teachers also understand density poorly, their discreet pieces of knowledge about density not being integrated into a coherent understanding. A number of factors contribute to density being a difficult concept. These include the presence of conceptions alternative to the particulate model of matter (Snir, Smith, & Raz, 2003), lack of distinction between the concepts of weight and density (Smith, Snir, & Grosslight, 1992), and between substances and objects (Wiser, Smith, Asbell-Clarke, & Doubler, 2009, April), difficulties with proportional reasoning (Rowell & Dawson, 1977), the complexity and abstractness of an intensive property which cannot be observed directly and which depends on the relationship between two variables (Kang, Scharmann, & Noh, 2004), and misunderstandings arising during teaching (Xu & Clarke, 2012). We considered the topic of density an appropriate one to use within an e-tutor for the PGCE NS course, given that density is generally poorly understood and incorporates multiple important scientific concepts relevant to the curriculum these students will need to be able to teach.

**Methodology**

**The case**

This is a convenience case study involving 50 student teachers who all have at least an undergraduate degree and who enrolled for the Natural Science (NS) methods course in the one year professional teacher education (PGCE) qualification at a university in South Africa. This qualifies them as generalist teachers of all subjects from grades 4 to 9. Case study research is well suited to gaining a rich understanding of a particular case, from which other practitioners and researchers can extract aspects relevant to their particular cases (Merriam, 1988).
The sample

The demographic information of all the participants in this case is summarised in Table 1.

Table 1. The students’ demographics and prior science education

<table>
<thead>
<tr>
<th>Number of students ((n)):</th>
<th>50</th>
<th>Female:</th>
<th>44</th>
<th>Male:</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age: (M = 26, SD = 6.1, \text{Range: 21 – 45} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students with prior teaching experience without a professional teaching qualification:</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students who completed science related school subjects:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 9 Natural science (compulsory):</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 12: Physical science (physics and chemistry):</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life science:</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earth science:</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students who completed science related courses at university level:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor of Science honours degree:</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemistry at year 1 level:</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master degree in Marine Biology:</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physics at year 1 level:</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students with no university science courses:</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students with Bachelor of Arts or Social Science degrees:</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As can be seen in Table 1, approximately half the students had exposure to physical science (PS), which includes both physics and chemistry, beyond the compulsory grade 9 school level. It is within PS that density and its related concepts are learnt, therefore the students’ prior exposure to PS is most relevant to this particular study. At least 36% (18 students) ascribed their choice not to have studied PS beyond the compulsory grade 9 level to negative attitudes toward PS, e.g., “I’ve always hated maths and science,” “My whole life I’ve hated science,” “It still petrifies me.” Remarks from students who had studied PS beyond grade 9 level regarding their experience of school PS include suggestions of: Low levels of enjoyment and long-term retention of the subject matter, e.g., “I didn’t remember any of my science: I hated it”; An algorithmic focus, e.g., “I just knew how to plug numbers into the formula and manipulate the formula, because that’s a lot of what we did in science”; Non-inclusive pedagogy e.g., “[I] sat at the back and no-one asked me anything,” “I used to ask lots of questions, and then if they didn’t answer, you just switch off, and it would be – like – really dry.” Nine of the students achieved an A or B symbol, i.e., above 70%, in the grade 12 PS exit level examination. Four studied PS at tertiary level. This discussion suggests a prevalent negativity toward PS as well as a large range in prior PS knowledge among the students.

![iQuiz](image)

Figure 2. The menu page of the density module
Implementation of the e-tutors in an integrated pedagogical approach

Class contact time for the density module was three hours a week for four weeks. In the first week the students were introduced to the module and answered a pre-test about density, and in the last they answered a post-test. Class-time engagement with the e-tutor occurred in the remaining two contact sessions. This involved working through nine e-tutors about various aspects related to density, as shown in the menu page (see Figure 2). Each student worked individually at a computer equipped with earphones. Both researchers were on hand to provide guidance on request. Students were instructed to discuss the work with one another if they wished, however they tended to work individually most of the time. The students were also introduced to a rich problem task at the start of the module, and required to answer this, in writing, for formal assessment, by the end of the module.

Data collection

Both quantitative and qualitative data were collected in a mixed method research design (Cresswell, 2003). The design was sequential explanatory Quant–Qual–Quant in which the qualitative data collected were used to elucidate and explain quantitative findings, and direct creation of later quantitative data collection instruments (McMillan & Schumacher, 2010).

Quantitative data and instruments

The quantitative data were collected in the forms of a pre-test, post-test, biographical survey and e-tutor engagement surveys of two types: A survey answered immediately after engagement with each of the nine e-tutors, and a survey answered at the end of the entire course. All of these can be found at www.angelastott.net/academic/appendices, with samples given in the Appendix. Fourteen of the 37 marks in the pre- and 100 marks in the post-tests came from questions developed, justified and motivated by Smith, Maclin, Grosslight, and Davis (1997) and seventeen marks from the 2003 TIMSS test (Mullis et al., 2005). All the items selected from these original standardised instruments had a high degree of construct validity.

The post-test consisted of all the pre-test questions, as well as additional non-transfer questions comprising 34 marks and transfer questions comprising 29 marks. Creation of these transfer questions was informed by guidelines given by Mayer (2009) and Dawkins et al. (2008).

Creation of the nine e-tutors was informed by Mayer (2009)’s principles of multimedia design, as well as research on effective conceptual teaching of density. This includes use of dot-representation diagrams, a qualitative, conceptual focus on the component concepts of matter, mass and volume and their relationship to density, as well as development of proportional reasoning before taking an algorithmic approach (Smith et al., 1992).

Responses were captured by the software as the students engaged with the e-tutors. This engagement data includes times spent and answers given. After engagement with each of the nine e-tutors the students answered six survey questions, each of which had quantitative and qualitative components. The quantitative components were seven-item Likert scale questions. Each item was given a description, with lower extents being assigned to lower item values. The survey questions were developed specifically for this research by the two researchers, guided by the research questions.

Qualitative data and instruments

An open-ended survey item formed the qualitative component of each survey question. Additionally, seven focus-group interviews were conducted and audio recorded. Interview protocols were informed by the survey questions. These further informed the design of the final evaluation survey with the aim of detecting the students’ overall e-tutor learning experience.
Data analysis

The quantitative data analysis included analysis of the participants’ pre- and post-test scores. Statistically significant differences in performance for the questions common between these two tests was determined by a p value < 0.05 for the ANOVA test performed.

The qualitative data analysis was done through an iterative process of engagement with, and manipulation of, the data guided by the research questions. This was done by the two researchers for data generated by the focus group interviews and open items in the respective surveys. An initial inter-rater agreement of 85% was achieved for the two raters on the qualitative category codes. Discrepancies between the interpretations of the two raters led to dialogue until consensus was reached. This guided subsequent analysis. Manipulation techniques used to aid analysis include summarisation of salient observations, coding of survey responses and focus-group transcriptions, and calculation and graphing of various indices from the quantitative data collected. Counts of the frequency of responses are given in brackets in the text as some indication of their relative weighting.

A multi-method type of triangulation design (Cohen, Manion, & Morrison, 2003) was used to generate data from various sources using multiple instruments to study the phenomena of deep learning and design features that enhance deep learning. The various data sources have been evaluated against one another by the two raters, who were the researchers, to enhance the validity of the findings.

Results

We now give a rich description organised according to our research questions and our theoretical framework.

Deep learning

To answer our first research question, regarding whether deep learning was evident as the students engaged with the module, we analysed the data for evidence of engagement in essential and generative processing, and self-regulated learning under intrinsic, mastery-oriented motivation. We also evaluated the extent to which students underwent extraneous cognitive processing, and the effect this had on learning depth.

Essential processing

As explained in the Mayer (2009) theoretical framework, essential processing can be detected by a learner’s performance on recall-type questions, i.e., those requiring responses similar to the presented information. Pre-post means for identical questions (Table 2) rose from 65% to 81%, with 40 of the 50 students making at least some improvement, 14 improving by over 25% and the mean improvement being 15%. This improvement was statistically significant, \( F(1,98) = 25.99, \ p < 0.05 \). It should be noted that the students received no feedback on the pre-test and had no access to either their answers or the question paper until after they had written the post-test, administered nine months after the pre-test. We therefore consider the pre-post score changes for the questions common between the two tests to be valid indicators of essential processing, rather than of superficial recall of answers due to prior exposure to the questions.

As can be seen in Figure 3, all 50 students obtained at least 50% for the non-transfer questions asked in the post-test, with 46 obtaining at least 70%. Some of these questions were identical to the pre-test and others were unique to the post-test but similar to the information which had been presented in the course. The mean post-test score obtained for these questions was 84%. These findings suggest that almost all the students did engage in considerable essential processing. This is necessary, but insufficient, for generative processing.
Generative processing

Generative processing, as explained in the theoretical framework, is synonymous to deep learning. The following comment suggests that heavy work-loads in the PGCE course tend to discourage generative processing: “I’m just like churning out the work: I don’t have any energy or time to actually do it properly.” However, all the students were able to score at least some marks on post-test transfer questions, with the mean score being 67%. Figure 3 shows the score distribution for the transfer questions, with those for the non-transfer questions included for comparison. For convenience, we define scores below 50% on transfer questions as evidence that low, 50% to 69% moderate, 70% to 79% high, and 80% to 100% very high, levels of generative processing had occurred. Based on this definition, 42 of the 50 students underwent moderate to very high levels of generative processing. This is consistent with the students’ self-reported seven-item Likert ratings regarding the extent to which they underwent deep learning. The mean rating of 4.5 falls between a moderate amount (12/42 students) and quite a lot (15).

Self-regulation

The students appear to have responded to the responsibility of deciding their own learning pace by making choices appropriate to their particular learning needs, i.e., by undergoing self-regulated learning. Evidence to support this includes the large ranges in the time-engagement data which seems to correspond to the large range in prior knowledge within the class. For example, the time to complete a single e-tutor ranges from a minimum of 30 seconds, for a student with good prior knowledge, to a maximum of 53.5 minutes during class and 124.3 minutes out of class time, for students with poor prior knowledge. Some students voluntarily spent considerable time engaging with the software. The course requirement was a single engagement with each of the nine e-tutors. However, only 14 of the 50 students did not repeat any of the e-tutors and the mean number of accesses per student was 15.3. One student repeated a single e-tutor nine times. Seven of the students repeated each of the nine e-tutors, but it was more common (15) for the students to only repeat those for topics they felt they had not yet mastered. Other evidence of self-regulation includes remarks (3) by stronger students that they self-directed their learning toward pedagogy since this, rather than content, was their individual learning need: “Not a deep learning in me per se, but an alternative way of viewing the concept of density from a learner’s mind frame and processing it from that point of view.”

Motivation

The test clearly provided a strong extrinsic motivation for the students to engage with the software beyond the minimum requirements of the course. This is evidenced by the spike in incidents of out of class e-tutor access in the
weekend (46 on Saturday, 132 on Sunday) and Monday morning (33) before the test. Nine of the students indicated
the test as their main motivation. Sixteen said that although this was initially their motivation, as they engaged with
the software a desire to master the concepts became a stronger motivation. Eight claimed they were driven primarily
by intrinsic mastery-oriented motivation throughout the module:

“I didn’t remember there was a test… I just saw it something we should learnt to teach. Because also, I’m
not that confident with science concepts and explaining them in a class, and having been in a school now
where I had to do a natural sciences lesson, it motivates you to try and understand.”

This student refers to learning for understanding (deep learning) being linked to a conceptual approach and intrinsic
motivation. The following student generalises the association between a conceptual approach and enjoyment:

“iQuiz you’re forced to think through each of the steps, and you’re forced to make sure you understand each
of the concepts before you can build on it. That’s good, because otherwise … I don’t think you’ll really
enjoy science if it was just like: Here’s a formula, use it, plug it in.”

One of the students referred to improvements in her levels of self-efficacy regarding knowledge of the content as
additional motivation for re-engagement with the e-tutors. During this re-engagement she found that she understood
the work and so progressed rapidly through the e-tutors with few errors, increasing her confidence and enjoyment: “I
[did] them all about five times. Even Sunday night, and I came in early … And I got it right, and it’s like: Yes!
Correct!” This student’s improvement in self-efficacy was not an isolated incidence: The pre-engagement ratings of
confidence to explain content had a 3.3 mean, where 1 referred to not at all confident (106 / 548 responses) and 7 to
extremely confident (14). The mean post-engagement rating was 4.7, with 406 responses indicating a rise in
confidence due to engagement with the software.

**Extraneous cognitive processing**

Extraneous processing is cognitive activity which impedes the learning process according to the Mayer (2009)
framework. Data used to evaluate the extent to which the students underwent extraneous cognitive processing
includes: A Likert-scale item and open-ended question, in the survey at the end of each e-tutor, about the extent to
which the student clicked without thinking and reasons for this; Examination of engagement data for students who
revisited slides, since this is an indication of the software having rerouted the student in circles, which may have
resulted from attempts to game the system.

Out of 546 survey responses regarding clicking without thinking, 332 were not at all, assigned a rating of 1, and
none were all / almost all the time, assigned 7. The mean rating, 1.9, corresponds closely to the option almost never
(82). Most (217/368) of the reasons given suggest a mastery-oriented epistemology favourable for deep learning. For
example:

“I’ve been through a lot of bad ones where to me it was just pointless and I went through it just click click
click. But when I started this immediately I realized… the information they’ve given me it’s actually
making me want to go in depth and really involve myself 100% in it. I think it’s … the way the program is
structured – it made me want to understand.”

Of the 368 responses about clicking without thinking: 35 referred to the intervention being easy. For some this was a
reason for clicking without thinking (22 responses): “I felt I knew the answers and was just going through the
motions of answering” and for some a reason for not clicking without thinking (13). Other explanations for clicking
without thinking include: Being in a hurry (9), boredom or tiredness (9), doing so by accident (8), lapses in
concentration (8), a false sense of knowing the content (5), technical and trialing issues (6), and not knowing how to
answer (22); “When I didn't know what the answer was I simply guessed and thus just clicked to get to the next
slide.” However, deep learning appears to have resulted despite this, in some cases, due to the repetitive and sub-step
designs of the software: “Because it still gives you the explanation of why it is correct,” “It’s really small steps so
even if you guess it’s almost as if it’s a progression to the next one.”

We considered 20 slide-revisits or fewer reasonable and so not indicative of an attempt to game the system. Almost
all (732/764) the times a student accessed an e-tutor, fewer than 20 slide revisits were recorded, with 343 involving
zero, and 269 one to five, revisits. The 32 cases of slide revisits greater than 20 were examined in detail. For 11 of these the students reported little to no attempts to game the system, and gave positive associated remarks. For example, a student who made 86 slide-revisits, spending 53 minutes on the e-tutor as a whole, remarked that it “actually challenged me to think critically and reason things out first before pressing the click button.” Even for the remaining 17 cases, in which students reported attempting to game the system, the majority gave positive ratings for experiencing the e-tutor as helpful (mean = 4.8/7) and enjoying it (mean = 5/7). A student who revisited 80 slides, spending 2.5 hours on the e-tutor, found it “excellent.” One student admitted trying to game the system for the first e-tutor, but reported that its design had deterred her from repeating this:

“Natural dislike for this subject so habit to try click through it to get it over! But it not letting me continue forced me to think about it.”

**Effective design features**

In the final evaluation 12 students responded to a question about design features which they found helpful in promoting deep learning. The features they identified were: Allowing them to work at their own pace (2), requiring them to think deeply (4), being visual, real and relevant (2), interactive (2), and interesting (2). During the focus group interviews we found additional evidence for each of these views. Some of this evidence is discussed below. Additionally, there was general agreement when some students referred to the value of the conceptual approach, as opposed to the very algorithmic approach they had experienced in school science, for example:

“I feel like in high school, yes, I remember the concept, but I feel like what I was given there was a recipe on how to calculate density or volume but I was never taught where it comes from – like the foundation. So now the iQuiz is like an eye-opener because now I understand volume, I understand density and I understand where it comes from.”

During the post-course focus group interviews there was also general agreement about the value of all the components of the module: The e-tutor, the rich problem task (*Archimedes task*), and the test, for promoting deep learning. For example, a student described how her initial engagement with the e-tutor led her to a basic understanding of the concepts so that she was able to understand the requirements of the *Archimedes task* as well as information she voluntarily accessed on the Internet. She was then able to use all this learning in answering the *Archimedes task*, developing understanding in the process:

“[For the task] you just have to pull [the concepts] together. So the process of engaging with the [task] really made me understand density… And it’s just that process of grappling with information to reach understanding.”

Only six of the 42 students who answered the final evaluation indicated that they would have preferred traditional lectures. Reasons for preferring the e-tutor to traditional lectures include being able to work at their own pace (17): “By doing it by yourself you can go at your own pace. If it is done in class and it is gone through slowly people could get bored,” enhanced interactivity (7): “We really got to engage with the work more so than we would have in lectures,” and the safety of answering without fear of ridicule (1): “It was very easy to understand and was done in a very neutral way with no fear of doing the wrong thing.” These responses were echoed in the focus group interviews by students with both extensive and limited science backgrounds:

“That was the biggest advantage for me: The fact that you could work through it at your own pace, because there were a couple of things I felt you could just go through easily and there were other things I knew if I needed to I could go back and redo it.”

“For the first time it wasn’t a proper chore”… “And I watched every video and wrote my notes.”…“I love that I can go at my own pace. And I thought: Imagine if you taught this in class. I wouldn’t have listened. I would have just done what everyone else does and I would never have understood and I could never have gone back and done it.”

The students were asked, in both survey questions and interviews, to comment on how the e-tutors’ sub-step design and branched structure affected their learning experience. Evidence for the contribution of each of these design features in guiding the learners toward deep learning has already been given. The majority of responses (436/542)
suggested that most students found the sub-step design effective in promoting deep learning. However, particularly students with higher prior content knowledge found the sub-step design tedious and frustrating (46 responses) and some (15 responses) ascribed their attempts to game the system to this. The branched and potentially repetitive structure of the e-tutors was a cause of frustration in some cases since a participant may be re-routed, possibly multiple times, to previously incorrectly answered questions: [I tried to game the system] “Because I got a bit tired of going around in circles and repeating some things.” This frustration and attempt to game the system were sometimes temporary and the repetition alerted the students to their need for learning, resulting in deeper learning. For example: “I think I went around thrice and I’m thinking: There’s something wrong here. And then I re-read … the question and I thought, okay but then it said this and this and that and then I sat and I stared at the computer and I was like really thinking and then it came to me after you know – wait a minute – it’s actually … A, and when I went ‘A’ and it was right, I was like ‘Oh that’s why!’ and then I rethought the process through.”

Conclusion, discussion, limitations and implications

We provide the following assertions in summary of the findings and discussion, and in answer to our research questions:

- Most of the students used the e-tutor engaged in a manner which suggests moderate to high levels of deep learning. This is suggested by evidence for limited extraneous processing associated with a lack of clarity, evidence for essential processing by most of the students, and evidence for generative processing, metacognitive control, and intrinsic mastery-oriented motivation among many of the students.
- Effective design features of the module include test and problem-based task inclusion, the sub-step design, branched structure, and conceptual approach of the electronic tutoring software.

In some cases the frustration which the students felt as the software rerouted them to previously incorrectly answered slides led to deep learning. This appears to be because the extraneous processing associated with this frustration made the student aware of their need to learn. This appears to contradict Mayer (2009)’s view of extraneous processing not serving an instructional goal. Instead, it appears to correspond to Schwartz et al. (2009)’s findings that problematisation of learning, associated with some extraneous processing, can motivate engagement in deep learning. We believe that extraneous cognitive processing which results from a lack of clarity does not aid deep learning. However, extraneous cognitive processing which results from discrepancy between the learners’ conceptual structures and the presented information, requiring the learner to recalibrate their self-assessment (Nicol & Macfarlane-Dick, 2006) and realise the need to undergo accommodation, does aid deep learning. It appears that most of the extraneous processing the students underwent as they engaged with the e-tutor was of the latter kind.

The value of the e-tutors’ sub-step design in promoting deep learning is consistent with VanLehn (2011)’s thesis that fine-grained formative feedback is crucial to a tutor’s success. The value of the conceptual approach to deep learning is consistent with Paul and Elder (2008)’s view that understanding is developed through conceptual, rather than algorithmic, engagement with subject matter. The value of inclusion of a rich problem task to accompany engagement with the e-tutors, for promoting deep learning, is consistent with work by Stott (2008) and Van Loggerenberg-Hattingh (2003).

As a convenience case study, the sample used is understood not to represent any particular population. It is possible that the PGCE students at the institution where the research was conducted are not similar to PGCE students in other institutions. These students’ access to the relevant resources for e-tutors to be effective are likely to be higher than for students in less developed parts of the world, particularly in the rest of Africa. This includes access to high speed broadband and computers both on campus and at home. Also, since the iQuiz e-tutor is conceptual in focus, it requires greater content and pedagogical knowledge, and skill and time to populate with appropriate content, than the more commonly available algorithmic ITSs (Graesser, Conley, & Olney, 2012). The resulting limitations to scalability may reduce the viability of the solutions presented here. However, as pointed out by VanLehn (2011), the reusability and effectiveness of such conceptual e-tutors may make their high initial costs worthwhile.

This study provides valuable insight into potential ways to improve teaching and learning in general using conceptual tutoring software to promote deep learning. It also adds to the sparse literature available on in-depth
mainly qualitative research about student engagement with, and effective design of, a somewhat extended learning module involving a conceptual e-tutor.

References


## Appendix

Sample questions from:

Pre-test (repeated in post-test)

<table>
<thead>
<tr>
<th>Question 1</th>
<th>Problem 1</th>
<th>Problem 2</th>
<th>Problem 3</th>
<th>Problem 4</th>
<th>Problem 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Look at the pairs of sugar-water mixtures alongside. Circle the correct option in each case.</td>
<td><img src="image1.png" alt="Problem 1" /></td>
<td><img src="image2.png" alt="Problem 2" /></td>
<td><img src="image3.png" alt="Problem 3" /></td>
<td><img src="image4.png" alt="Problem 4" /></td>
<td><img src="image5.png" alt="Problem 5" /></td>
</tr>
<tr>
<td>Problem 1: The sweeter mixture is:</td>
<td>Problem 1: The sweeter mixture is:</td>
<td>Problem 2:</td>
<td>Problem 3: The sweeter mixture is:</td>
<td>Problem 4: The sweeter mixture is:</td>
<td>Problem 5: The sweeter mixture is:</td>
</tr>
<tr>
<td>[A / B / both A and B are the same sweetness]</td>
<td>[A / B / both A and B are the same sweetness]</td>
<td>[A / B / both A and B are the same sweetness]</td>
<td>[A / B / both A and B are the same sweetness]</td>
<td>[A / B / both A and B are the same sweetness]</td>
<td>[A / B / both A and B are the same sweetness]</td>
</tr>
</tbody>
</table>

QUESTION 17

You are a Natural Sciences teacher. You are teaching your class about density. Explain how you would answer a learner who makes each of the following statements. In your answer you need to say whether the learner’s statement is correct or not, and why this is so. This must be done in a way that helps all the learners in the class to understand density better. You can make use of examples and/or pictures to help make your explanation clearer.

17.1. “It is possible for a big object to be lighter than a smaller object. “

17.2. “Block A is 1 kg. Block B is 0.5 kg. If Block A floats, B will definitely float too.”

17.3. “For the diagram alongside, B is denser than A because B’s mass is squashed into a smaller space than A’s. But this might not mean that B’s particles are squashed together more than A’s.”

---

Biographical survey

Permission to collect data

If you give permission for us to use this data, this will be anonymous and every effort will be made that no harm will come to you from this. Your name is only asked in the questionnaire for logistical purposes. It will never appear in any publication. Do you give us permission to use this data?

☐ yes  ☐ no

Yourself

Surname
Name
Age

Gender:
☐ male  ☐ female

School science learning

Did you complete grade 12 in South Africa?
☐ yes
Did you take natural science up to grade 9?
- yes
- no

Give your approximate grade 9 natural science mark
- A
- B
- C
- D
- E or below

**e-tutor engagement survey answered after each e-tutor engagement**

**Helpfulness**

How helpful was this tutoring session?
- 1 A waste of time
- 2 Hardly helpful
- 3 Slightly helpful
- 4 Fairly helpful
- 5 Helpful
- 6 Very helpful

Why was this tutoring session helpful / not helpful?

**Meaningless clicking**

How much did you click links without thinking?
- 1 Not at all
- 2 Almost never
- 3 A few times
- 4 A moderate number of times
- 5 Fairly many times
- 6 A lot
- 7 All / almost all of the time

Why did you / did you not click without thinking?

**Final e-tutor engagement survey**

**How you engaged with iQuiz**

Did you take notes while you engaged with iQuiz?
- 1 Never
- 2 I took a few very short notes
- 3 I took notes for a fair amount of the work
- 4 I took notes for much of the work
- 5 I took notes for all of the work
Why did you / did you not take notes while engaging with iQuiz?

What was your main motivation for engaging with the density iQuizzes?
- [ ] 1 Having to tick off the list
- [ ] 2 The test
- [ ] 3 Wanting to learn the knowledge
- [ ] 4 Initially the test / ticking the list and later wanting to learn the knowledge
- [ ] 5 Other
If you say ‘other’, please say what this was

Why was this your main motivation for engaging with iQuiz?

Did you repeat any of the iQuizzes?
- [ ] 1 None
- [ ] 2 Yes, I randomly repeated some
- [ ] 3 Yes, I purposefully repeated those on topics which I felt I needed to work more on
- [ ] 4 Yes, I repeated all of them
ABSTRACT
The present study investigates empirical models of groupwork management in online collaborative learning environments, based on the data from 298 students (86 groups) in United States. Data revealed that, at the group level, groupwork management was positively associated with feedback and help seeking. Data further revealed that, at the individual level, groupwork management was positively associated with feedback, peer- and learning-oriented reasons, help seeking, and the number of online courses. In addition, older students were found more frequently to manage online groupwork. The findings are discussed in the context of related literature in the field. Our study suggests directions for practice and future research regarding online groupwork management.

Keyword
Collaborative learning, Homework, Online groupwork, Self-regulated learning

Introduction
The trend towards online learning has propelled many instructors to incorporate online groups in their classes. Online groups are usually small (e.g., ranging from two to four students) and heterogeneous groups, designed to help online students, for example, develop problem-solving skills, share and challenge one another’s ideas, and better prepare them for future careers (Jonassen, 2000; Koh, Barbour, & Hill, 2010; Smith et al., 2011).

Yet, online groupwork presents new challenges for students. They are required to manage online groupwork, including arranging their online and offline study environments (Deimann & Bastiaens, 2010; Whipp & Chiarelli, 2004), coordinating time for groupwork (Biasutti, 2011; Gafni & Geri, 2010), handling online and offline distractions (Bigenho, 2011; Whipp & Chiarelli, 2004), keeping themselves motivated (Liu, Joy, & Griffiths, 2010; Smith et al., 2011), and coping with negative emotions in the groupwork process (Ku, Tseng, & Akarasriworn, 2013; Wosnitza & Volet, 2005).

As a number of scholars (Choi & Kang, 2010; Rovai, 2007; Volet, Vauras, & Salonen, 2009) have been calling for attention to these challenges, it is obvious that these issues warrant our investigation. Thus, it would be important to examine factors that may influence students’ groupwork management in online environments. A study such as this is especially important, as students tend to hold a more negative attitude toward online groupwork than face-to-face groupwork (e.g., Smith et al., 2011), and as they often face additional logistical issues (e.g., varying time zones and work schedules, and fewer channels for communication) in engaging in group-oriented online activities (Havard, Du, & Xu, 2008; Liu et al., 2010).

Theoretical framework
One framework pertaining to online groupwork management is a self-regulated learning perspective (Pintrich, 2004; Zimmerman, 2008), particularly from the view of volitional control (Corno, 1993, 2004; Deimann & Bastiaens, 2010). The term volition refers to both the strength of will to complete a task and the diligence of pursuit (Corno, 1993). It is primarily concerned with issues of implementation that occur after a specific goal is set, to sustain one’s intention to follow through the goal despite temptations or competing personal strivings. It involves a range of
regulatory activities of a purposeful striving toward the attainment of the goal, including, for instance, budgeting time, monitoring motivation, and controlling emotion (Corno, 2004).

Volitional control is particularly critical to managing online groupwork because goals of groupwork are usually set by online instructors; the main task for students is to deal with the demands of following through online groupwork. They are required to sustain the needed focus and effort to complete their work in online collaborative learning environments, with less structure and time constraints than exist in face-to-face classrooms.

A self-regulation perspective (Garcia, McCann, Turner, & Roska, 1998; Pintrich & Zusho, 2002; Schunk, 2005) suggests that the use of self-regulatory strategies is influenced by goal orientation (e.g., purpose), task value (e.g., importance), and task interest (e.g., appeal). For example, Garcia et al. (1998) linked volition with expectancy-value theory (Eccles, 1983), particularly task value to implementing and protecting intention. They further hypothesized that volitional control is affected by the pleasure an individual experiences while engaging in a specific task and the value one attaches to the task. Specifically, based on expectancy-value theory, Warton (2001) pointed to the importance of task value in completing a task, including (a) task interest and (b) task utility (the purposes of a task).

As an individual with a greater interest in a task and viewing it as more important is more likely to employ self-regulatory strategies (Pintrich & Zusho, 2002), interest and task value may influence self-regulation (Schunk, 2005), with online groupwork management in particular. Furthermore, as a self-regulation perspective recognizes that individual and environmental differences may influence self-regulation (Pintrich, 2004), online groupwork management may be affected by student characteristics as well as help seeking and feedback. For example, seeking help from peers may enhance a students’ efforts to manage their work (Wolters, 2011), which may further affect peers in online collaborative learning environments.

Taken together, this body of literature suggests that online groupwork management may be affected by a range of variables such as goal orientation, task value, task interest, affective attitude, help seeking, feedback, and student characteristics. Consequently, there is a need to incorporate these variables in models of groupwork management.

Studies pertaining to online groupwork management

As more university instructors incorporate online groupwork in their courses, a growing number of studies have examined student experiences with this (Gafni & Geri, 2010; Rovai, 2007; Smith et al., 2011; Wosnitza & Volet, 2005). For example, Smith et al. (2011) found that students held a more negative attitude toward online groupwork than face-to-face groupwork (e.g., less motivated to engage in groupwork). More specifically, Gafni and Geri (2010) found that students were more likely to follow through their individual task on time, whereas they tended to put off the collaborative aspect of the task until the final weeks of the semester. Similarly, Wosnitza and Volet (2005) found that online students often faced the issue of social emotion in online collaborative group settings. For example, one student was “really angry” because one group member was hardly made any contribution to online groupwork, which raised the workload for other group members.

Still other studies alluded to the emerging issue of environmental control in online learning environments. Whipp and Chiarelli (2004) examined students’ self-regulatory strategies in an online technology course. Data revealed that students used different self-regulated learning strategies (e.g., setting up quiet areas in their homes for their online work). Their study implied that the use of self-regulatory strategies was affected by motivational influences (e.g., goal orientation) and environmental influences (e.g., helpful feedback from instructor). It further implied that the online groupwork may present additional challenges for group regulation, as what is at stake for groupwork is not just about managing one’s own environment. More important, it is also about managing their group members’ study environment (i.e., due to increased interdependence on peers).

To summarize, this line of studies suggests that groupwork presents multiple challenges for group members in online environments. However, few studies have examined multiple factors that may influence students’ effort at managing these challenges.
The present investigation

The aim of the present investigation was to link students’ groupwork management in online environments to multiple variables at the individual and group level. Understanding of these relevant variables and factors for online groupwork management will provide new insights about how to create a better learning environment for online collaborative groupwork. Our empirical investigation was carried out in the context of multi-level modeling, to address the issues introduced by nested data structure (i.e., individuals nested under groups in the online collaborative learning groupwork setting). Several related multi-level models were examined, concerning the particular predictor variables included and the level of these variables.

Method

Participants

The participants in the present study consisted of 298 students (86 groups) from one university in the southeastern United States. These online groups ranged from 2 to 4 students ($M = 3.47, SD = .63$). Specifically, the number of the groups with 2, 3, and 4 students were 6 (7.0%), 34 (39.5%), and 46 (53.5%), respectively.

Of these participants, 167 were females (56.0%) and 131 were males (44.0%). The racial/ethnic breakdown was 139 Caucasians (46.6%), 138 African Americans (46.3%), and 21 students from other backgrounds (7.1%). Specifically, the number of all-male groups, all-female groups, and mixed gender groups were 12 (14.0%), 17 (19.8%), and 57 (66.3%). Meanwhile, the number of all-Caucasian groups, all-African-American groups, and mixed racial/ethnic groups were 14 (16.3%), 14 (16.3%), and 58 (67.4%).

Approximately three-fourths of the participants were full-time students (76.5%). Two-thirds of the participants were 30 years or younger (67.2%), whereas one-third of them were over 30 years old (32.8%). These participants were recruited from one graduate-level course from 2009 to 2012. No significant differences were found across these years relating to their background characteristics (e.g., race, age, and gender).

Online groupwork

The course focused on theories and principles of instructional design, strategies for developing multimedia and for applying design models, and evaluation of related educational software. It was delivered through myCourses, which was used to deliver online courses through various communication tools such as chat rooms, discussion boards, and emails.

In the beginning of the semester, the participants were randomly assigned to online groups by the instructor. Each group was then asked to select a group leader, who was responsible for facilitating the flow of the groupwork (e.g., coordinating specific tasks among the group members such as designing the website, incorporating relevant images, and writing up the paper). The final group project required each group to select a genuine instructional problem, to design and develop a plan to solve this problem, and to present its evaluation. The participants were required to involve various discussion activities with their members through myCourses.

Guidance and training on group collaboration were provided throughout the project planning, execution, and reflection process. When planning the project, the participants received written guidelines with evaluation criteria for individuals and groups. Training was then offered for each group leader in online chat rooms relating to group function. In addition, each group was asked to present its proposal with required contents including division of labor and time management. Advice was then solicited from peers and the instructor. Following that, online tutoring was offered to tackle specific problems encountered by respective groups to them to timely adjust group collaboration in the process. The participants were further asked to include in self-reflection in their presentation on how they approached collaboration.
Measures

Students were asked about how many online courses they took previously, ranging from none to four or more. A number of scales were used in the current investigation (see Table 1).

Table 1. Scales and sample items

<table>
<thead>
<tr>
<th>Scales (Number of items)</th>
<th>Sample items</th>
<th>α (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback a(5)</td>
<td>Monitored by the group members. Monitored by the instructor.</td>
<td>.80 (.77, .84)</td>
</tr>
<tr>
<td>Peer-oriented reasons b(4)</td>
<td>Participating online GW brings you approval from group members. Participating online GW gives you opportunities to work with group members.</td>
<td>.78 (.74, .82)</td>
</tr>
<tr>
<td>Learning-oriented reasons b(5)</td>
<td>Participating online GW helps you learn communication skills. Participating online GW helps you work more productively.</td>
<td>.83 (.80, .86)</td>
</tr>
<tr>
<td>Online GW interest c(5)</td>
<td>I look forward to online GW. Online GW is fun. I enjoy online GW.</td>
<td>.94 (.93, .95)</td>
</tr>
<tr>
<td>Affective attitude (4)</td>
<td>My motivation or desire to participate online GW is _____ d(online individual work) My attention while participating online GW is _____ d(online individual work) My moods while participating online GW is _____ e(online individual work)</td>
<td>.86 (.83, .88)</td>
</tr>
<tr>
<td>Help seeking f(7)</td>
<td>I ask the instructor to clarify concepts I don’t understand well. I try to identify students in my group whom I can ask for help if necessary. I try to identify online resources where I can get help if necessary.</td>
<td>.82 (.79, .85)</td>
</tr>
<tr>
<td>Online GW management g(34)</td>
<td>Find a quiet place Keep track of what remains to be done Plan ahead Stop GW to surf the Internet h Tell my group members to calm down</td>
<td>.92 (.90, .93)</td>
</tr>
</tbody>
</table>

Note. GW = groupwork. All the ratings were based on student survey.

The 95% confidence intervals (CI) for coefficient alpha were calculated using a method employing the central F distribution (see Fan & Thompson, 2001).

a Rating: 1 (none), 2 (some), 3 (about half), 4 (most), and 5 (all).
b Rating: 1 (strongly disagree), 2 (disagree), 3 (agree), and 4 (strongly agree).
c Rating: 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), and 5 (strongly agree).
d Rating: 1 (much lower than), 2 (lower than), 3 (about the same as), 4 (higher than), and 5 (much higher than).
e Rating: 1 (much worse than), 2 (worse than), 3 (about the same as), 4 (better than), and 5 (much better than).
f Rating: 1 (not at all true of me) to 7 (very true of me).
g Rating: 1 (never), 2 (rarely), 3 (sometimes), 4 (often), and 5 (routinely).
h The item was reverse scored.

Feedback. Based on relevant literature (Trautwein, Koller, Schmitz, & Baumert, 2002; Xu, 2008a), this scale consisted of five items to measure how much groupwork was monitored by group members and the instructor (α = .80; the amount of the groupwork being checked and shared).

Reasons for online groupwork. This scale consisted of two subscales (peer- and learning-oriented reasons), informed by the Homework Purpose Scale (Saban, 2013; Xu, 2010, 2011). Five items assessed learning-oriented reasons (α = .83), with respect to media, technology, and productivity. Four items assessed peer-oriented reasons (α = .78), regarding working and collaborating with group members. Confirmatory factor analysis indicated that this scale consisted of these two related but separate subscales (SRMR = .046; RMSEA =.045; 90% CI = .013, .070; CFI = .973).
Online groupwork interest. This scale consisted of five items to measure students' interest toward online groupwork ($\alpha = .94$), based on related literature on interest and intrinsic motivation (Wigfield & Eccles, 2000; Xu, 2008a; Xu & Corno, 1998). It measured the extent to which students look forward to online groupwork and to what extent they like groupwork assignments.

Affective attitude. Informed by related literature (Warton, 2001; Xu, 2008a), this scale consisted of four items to measure the favorability of online groupwork (e.g., students’ motivation, attention, and moods; $\alpha = .86$).

Help seeking. This scale incorporated seven items to measure students’ effort for help seeking in the online groupwork process ($\alpha = .82$), informed by and adapted from relevant items from the Motivated Strategies for Learning Questionnaire (Duncan & McKeachie, 2005).

Online groupwork management. As the outcome variable of our research in this study, this scale was developed from related literature (Corno, 2004; Corno & Xu, 2004; Xu, 2008b, 2008c; Xu & Corno, 2003). It consisted of 34 items ($\alpha = .92$), including managing time, handling distraction, arranging the environment, controlling emotion, and monitoring motivation (see sample items in Table 1).

Statistical modeling analyses

As we had nested data structure (individuals nested under groups), we used multi-level modeling analysis to appropriately take care of a number of major issues (e.g., heterogeneity of regression and misestimated standard errors) introduced by the nested data structure. Multilevel modeling can include variables at different levels, by taking into consideration the nonindependence of observations (Raudenbush & Bryk, 2002).

In the current study, multilevel analyses were performed using the HLM 6. To help interpret the resulting regression coefficients, we first standardized all continuous variables ($M = 0, SD = 1$) prior to conducting the multilevel analyses. As a result, the regression weights for all variables (except the dummy-coded variables) were largely equivalent with the standardized weights from multiple-regression procedures (Xu, 2008a).

Model 1 included ten student-level variables, while no group-level predictors were used. These variables included student characteristics (gender, age, previous online courses, and student status), feedback, peer- and learning-oriented reasons, online groupwork interest, affective attitude, and help seeking. We hypothesized that groupwork management was positively associated with feedback, peer- and learning-oriented reasons, online groupwork interest, helping seeking, and affective attitude. These hypotheses were informed by self-regulation literature that students' management in groupwork may be affected by task value, task interest, affective attitude, feedback, and help seeking (Garcia et al., 1998; Pintrich, 2004; Pintrich & Zusho, 2002; Schunk, 2005). In addition, these hypotheses were consistent with empirical findings that secondary students’ homework management was positively related to feedback, homework purpose, homework interest, and affective attitude toward homework (Xu & Wu, 2013). Student characteristics (e.g., age and gender) were included in the model for statistically controlling for these characteristics in our study.

Model 2 incorporated additional four variables at the group level (i.e., online groupwork interest, affective attitude, feedback, and help seeking). Each group-level variable was based on the aggregation of the variable within a group to the group level to build an index of the shared perception (e.g., shared groupwork interest). The main reason for including these group-level variables was based on the consideration that the use of regulation strategies could be influenced by academic and social environments (Corno & Mandinach, 2004), including values, norms, and expectations concerning feedback, help seeking, and affective engagement in online groupwork. For example, students’ shared feedback in a given group could have an effect on online groupwork management over and above the effect of feedback at the individual level.

Our models implemented in the present study were random-intercept models (Raudenbush & Bryk, 2002), in that the intercept of these models was considered random to reflect between-group differences in online groupwork management. We did not estimate the random parts of the slopes, since we had no a priori hypotheses regarding between-group differences in the predictive power of the predictor variables. Full maximum likelihood was applied in all models. We centered four group-level variables (help seeking, feedback, online groupwork interest, and
affective attitude) at the group mean, in order to separate compositional and individual effects (Raudenbush & Bryk, 2002).

**Results**

Table 2 includes the descriptive statistics with respect to all independent variables and the outcome variable of groupwork management. In addition, it includes the Pearson correlations among all of the study variables. Groupwork management was significantly related to all of the independent variables.

![Table 2. Descriptive statistics and Pearson correlations](image)

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gender (male: 1)</td>
<td>.44</td>
<td>.50</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Full-time student (yes: 1)</td>
<td>.77</td>
<td>.42</td>
<td>.36†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Age (&lt;=30: 0, &gt; 30: 1)</td>
<td>.33</td>
<td>.47</td>
<td>.35†</td>
<td>-.50†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Previous online courses</td>
<td>2.27</td>
<td>1.55</td>
<td>-.29†</td>
<td>-.21†</td>
<td>.17†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Feedback</td>
<td>3.25</td>
<td>.90</td>
<td>-.27†</td>
<td>-.23†</td>
<td>.23†</td>
<td>.27†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Peer-oriented reasons</td>
<td>2.97</td>
<td>.53</td>
<td>-.13</td>
<td>-.13</td>
<td>.22†</td>
<td>.20†</td>
<td>.51†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Learning-oriented reasons</td>
<td>2.91</td>
<td>.60</td>
<td>-.10</td>
<td>-.03</td>
<td>.23δ</td>
<td>.45†</td>
<td>.63†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Online GW interest</td>
<td>3.11</td>
<td>.97</td>
<td>-.14</td>
<td>-.15†</td>
<td>.07</td>
<td>.20†</td>
<td>.36†</td>
<td>.36†</td>
<td>.49†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Affective attitude</td>
<td>2.98</td>
<td>.80</td>
<td>-.10</td>
<td>-.20†</td>
<td>.06</td>
<td>.05</td>
<td>.25†</td>
<td>.36†</td>
<td>.42†</td>
<td>.78†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Help seeking</td>
<td>4.91</td>
<td>1.04</td>
<td>-.03</td>
<td>-.09</td>
<td>-.01</td>
<td>-.05</td>
<td>.14†</td>
<td>.26†</td>
<td>.23†</td>
<td>.23†</td>
<td>.31†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Feedback (group)</td>
<td>3.28</td>
<td>.61</td>
<td>-.21†</td>
<td>-.25†</td>
<td>.26†</td>
<td>.38†</td>
<td>.66†</td>
<td>.41†</td>
<td>.35†</td>
<td>.25†</td>
<td>.17†</td>
<td>.10†</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Online GW interest (group)</td>
<td>3.13</td>
<td>.59</td>
<td>-.10</td>
<td>-.15†</td>
<td>.06</td>
<td>.30†</td>
<td>.27†</td>
<td>.17†</td>
<td>.25†</td>
<td>.60†</td>
<td>.46†</td>
<td>.07</td>
<td>.42†</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Affective attitude (group)</td>
<td>3.00</td>
<td>.51</td>
<td>-.11</td>
<td>-.17†</td>
<td>.09</td>
<td>.14†</td>
<td>.20†</td>
<td>.25†</td>
<td>.24†</td>
<td>.45†</td>
<td>.61†</td>
<td>.16†</td>
<td>.31†</td>
<td>.75†</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>14. Help seeking (group)</td>
<td>4.91</td>
<td>.63</td>
<td>-.06</td>
<td>-.14†</td>
<td>.03</td>
<td>-.02</td>
<td>.11</td>
<td>.25†</td>
<td>.13†</td>
<td>.09</td>
<td>.18†</td>
<td>.58†</td>
<td>.17†</td>
<td>.14†</td>
<td>.30†</td>
<td>---</td>
</tr>
<tr>
<td>15. Online GW management</td>
<td>3.58</td>
<td>.55</td>
<td>-.21†</td>
<td>-.12†</td>
<td>.23†</td>
<td>.28†</td>
<td>.55†</td>
<td>.58†</td>
<td>.52†</td>
<td>.35†</td>
<td>.28†</td>
<td>.37†</td>
<td>.38†</td>
<td>.16†</td>
<td>.14†</td>
<td>.26†</td>
</tr>
</tbody>
</table>

**Note.** GW = groupwork. N = 298.

The aggregation of the variable within a group to form an index of the shared view at the group level (e.g., shared feedback). The fully unconditional model showed that most of the variance occurred at the individual level, with 14.0% of the variance in online groupwork management occurred at the group level. For this model, the deviance statistics and the associated number of estimated parameters were 837.98 and 3. \( p < .05. \) \( \dagger p < .01. \) Model 1 with ten student-level variables had deviance statistics and the associated number of estimated parameters being 639.53 and 13. We used the likelihood ratio test to compare Model 1 (i.e., with ten student-level-variables) to the fully unconditional model (i.e., without ten student-level-variables). As hypothesized, the result showed that Model 1 provided a statistically significantly better fit than the fully unconditional model (\( \chi^2 \text{ (df = 10) } = 198.45; p < .001). Model 1 (i.e., with ten student-level variables) explained 43.88% of the variance in groupwork management at the individual level, and explained 72.88% of the variance in groupwork management at the group level (see Table 3).

Model 2 consisted of four group-level variables (feedback, help seeking, online groupwork interest, and affective attitude). For this model, the deviance statistics and the associated number of estimated parameters were 612.66 and 17. We used the likelihood ratio test to compare Model 2 (i.e., with four-group level variables) to Model 1 (i.e., without four-group level variables). As hypothesized, the result showed that Model 2 provided a statistically significantly better fit than Model 1 (\( \chi^2 \text{ (df = 4) } = 26.87, p < .001). Data further indicated that Model 2 explained an additional 1.22% of the variance in online groupwork management at the individual level as well as an additional 27.08% of the variance in online groupwork management at the group level.

Taken together, Model 2 accounted for 45.10% of the student-level variance in groupwork management, 99.96% of the group-level variance, and 52.79% of the total variance. As indicated in Table 3, groupwork management was positively associated with feedback (\( b = .32, p < .01), peer-oriented reasons (\( b = .22, p < .01), help seeking (\( b = .21, p < .01), learning-oriented reasons (\( b = .17, p < .05), and the number of previous online courses (\( b = .15, p < .01). In addition, those students who were over 30 years old (compared with those students who were 30 and below) reported
more frequently working to manage online groupwork \( (b = .30, p < .05) \). Finally, groupwork management was positively associated with two group-level variables: help seeking \( (b = .32, p < .01) \) and feedback \( (b = .23, p < .05) \).

### Table 3. Online groupwork management: Results from multilevel modeling

<table>
<thead>
<tr>
<th>Model predictor</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b )</td>
<td>( SE )</td>
<td>( b )</td>
<td>( SE )</td>
</tr>
<tr>
<td><strong>Student level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (female: 0, male: 1)</td>
<td>-.07</td>
<td>.12</td>
<td>-.04</td>
<td>.11</td>
</tr>
<tr>
<td>Full-time student (no: 0, yes: 1)</td>
<td>.13</td>
<td>.14</td>
<td>.23</td>
<td>.14</td>
</tr>
<tr>
<td>Age ( (&lt;=30: 0, &gt;30: 1) )</td>
<td>.30*</td>
<td>.13</td>
<td>.30*</td>
<td>.13</td>
</tr>
<tr>
<td>N of previous online courses</td>
<td>.16†</td>
<td>.05</td>
<td>.15†</td>
<td>.04</td>
</tr>
<tr>
<td>Feedback</td>
<td>.29†</td>
<td>.07</td>
<td>.32†</td>
<td>.08</td>
</tr>
<tr>
<td>Peer-oriented reasons</td>
<td>.28†</td>
<td>.06</td>
<td>.22†</td>
<td>.06</td>
</tr>
<tr>
<td>Learning-oriented reasons</td>
<td>.19†</td>
<td>.07</td>
<td>.17†</td>
<td>.07</td>
</tr>
<tr>
<td>Online groupwork interest</td>
<td>.01</td>
<td>.08</td>
<td>.03</td>
<td>.08</td>
</tr>
<tr>
<td>Affective attitude</td>
<td>.03</td>
<td>.09</td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td>Help seeking</td>
<td>.21†</td>
<td>.06</td>
<td>.21†</td>
<td>.06</td>
</tr>
<tr>
<td><strong>Group level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
<td></td>
<td>.23*</td>
<td>.09</td>
</tr>
<tr>
<td>Online groupwork interest</td>
<td></td>
<td></td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td>Affective attitude</td>
<td></td>
<td></td>
<td>-.17</td>
<td>.10</td>
</tr>
<tr>
<td>Help seeking</td>
<td></td>
<td></td>
<td>.32†</td>
<td>.08</td>
</tr>
<tr>
<td>Variance explained at the student level</td>
<td>43.88%</td>
<td></td>
<td>45.10%</td>
<td></td>
</tr>
<tr>
<td>Variance explained at the group level</td>
<td>72.88%</td>
<td></td>
<td>99.96%</td>
<td></td>
</tr>
<tr>
<td>Total variance explained</td>
<td>47.95%</td>
<td></td>
<td>52.79%</td>
<td></td>
</tr>
<tr>
<td>Model deviance</td>
<td>639.53</td>
<td></td>
<td>612.66</td>
<td></td>
</tr>
<tr>
<td>Number of estimated parameters</td>
<td>13</td>
<td></td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

Note. \( N = 298 \) students from 86 online groups. \( b \) = unstandardized regression coefficient. \( SE \) = standard error of \( b \). Likelihood-ratio tests that compare models can be calculated when one compares differences in model deviance values with degrees of freedom equal to differences in number of estimated parameters. \* \( p < .05 \). \† \( p < .01 \).

### Discussion

Our study examined empirical models of variables to predict groupwork management in online collaborative learning environments. Data indicated that six student-level variables explained the variance in online groupwork management (i.e., age, previous online courses, peer- and learning-oriented reasons, feedback, and help seeking). At the group level, online groupwork management was positively associated with feedback and help seeking.

How do we explain the finding that groupwork management was positively associated with previous online courses? Whereas Barnard-Brak, Paton, and Lan (2010) did not find any difference in self-regulatory skills (e.g., environmental structuring and time management) over students’ first semester of online learning, they hypothesized that one semester may not be long enough to assess significance difference in this area. While in line with their hypothesis, our study further suggests that it may take time for self-regulatory strategies to develop in online learning environments.

The finding relating to age difference is consistent with previous finding (Cooper & Corpus 2009) that adults (ranging in age from 18 to 22) demonstrated more knowledge of the effectiveness of strategies for sustaining motivation that did fifth graders. Thus, the present study expands previous research, by suggesting that older adults (ranging from 31 to 60) are more likely to manage groupwork than do younger adults (being 30 years or below) in online collaborative learning environments.

These findings (i.e., relating to the differences by age and online courses) suggest that instructors need to pay special attention to those students who are younger and with less online learning experience, to help them to better manage online groupwork. This may include (a) sharing effective online groupwork management strategies (e.g., arranging a
Our findings that groupwork management was positively related to help seeking and feedback are consistent with the self-regulation literature (Corno, 2004; Pintrich, 2004; Wolters, 2011) and previous findings on students’ homework management (Xu & Wu, 2013). In addition, our study extends previous research in the field, by revealing that help seeking and feedback in a given group had positive effects on groupwork management in addition to their positive effects at the individual level. These findings suggest that feedback and help seeking provide students with added incentive and willingness (in the constant presence of the supportive instructor and group members) to exert efforts to regulate their groupwork in online collaborative learning settings.

A recent literature review on formative feedback suggests that uncertainty can interfere with task performance, whereas decreasing uncertainty through formative feedback can enhance motivation and contribute to more effective task management strategies (Shute, 2008). As ones activities are less visible to peers in online environments (An, Kim, & Kim, 2008), a student may experience a higher level of uncertainty in the online groupwork process. Thus, it is not surprising that feedback has a more important role in leading to effective task management strategies, with online groupwork management strategies in particular (e.g., accessing group members and organizing around their schedule). This is, to some extent, substantiated by recent qualitative findings regarding the importance of feedback (e.g., sharing and expanding relevant ideas) in following through online group projects (Biasutti & El-Deghaidy, 2012; Ku et al., 2013). Thus, it would be desirable to promote a norm of providing ongoing feedback among the instructor and group members, such as providing constructive comments of each others’ progresses and offering timely suggestions to prevent group members from going off course.

Similarly, it would be beneficial to develop and foster a norm of help seeking, to encourage online students to seek help from various sources (e.g., group members and online sources) through various channels (e.g., discussion board and web chat), ranging from clarifying groupwork expectations to offering suggestions about how to better approach certain aspects of groupwork. Furthermore, as online students may use Web-based helpers as well as peers’ online discussions and submissions to plan their work and monitor their progress (Whipp & Chiarelli, 2004), it would be beneficial to promote a norm of encouraging students to take advantage of this unique form of help seeking in Web-based learning environments, a form of help seeking that is more readily accessible and relatively less obtrusive.

Consistent with the theoretical claim regarding the important role of goals and task importance on self-regulation (Pintrich, 2004), our study takes another step forward, by showing that peer- and learning-oriented reasons were positively associated with online groupwork management. Thus, our study extends previous research concerning the importance of goals in self-regulation from individual-oriented tasks in face-to-face environments (e.g., Pintrich & Zusho, 2002; Xu & Wu, 2013) to group-oriented tasks in online learning environments. Consequently, it would be helpful for instructors to develop online collaborative activities to engage online students. Furthermore, as online groupwork requires positive interdependence among group members (Nam & Zellner, 2011), it would be valuable to foster group cohesiveness in the progress. It would also be helpful to promote a norm of sharing exemplary groupwork management strategies with peers, so that they can learn with and from each other about how to better manage the challenges associated with online groupwork.

Meanwhile, how do we make out of the findings that online groupwork management was not associated with interest and affective attitude toward online groupwork? These findings are not in line with (a) the theoretical claim regarding the role of task interest on self-regulation (Garcia et al., 1998; Pintrich & Zusho, 2002) and (b) previous findings relating to secondary students’ homework management (Xu & Wu, 2013). One possible explanation is that, for graduate students in our study, task importance (compared with task interest) may play a more central role in their online collaborative learning environments. This is an important hypothesis to be examined in future investigation. This hypothesis is, to some extent, substantiated by the observation that children’s involvement in different activities may be based more on their interest in the activity rather than its usefulness (Wigfield, Tonks, & Eccles, 2004). It is further consistent with the findings that, for undergraduate students, task-specific importance (compared with task-specific interest) played a more significant role in test-taking effort (Cole, Bergin, & Whittaker, 2008).

It is important to state that our findings were derived from a relatively large sample of students, using multilevel analyses. On the other hand, it has some limitations that need to be acknowledged. One limitation is that the present study was based on self-report. Another limitation is that our research design was correlational. Whereas feedback,
help seeking, peer- and learning oriented reasons were found to predict online groupwork management, these variables were not manipulated. Consequently, it would be informative for future research to manipulate some of these variables and to examine the influences of these manipulations on subsequent online groupwork management.

It would be important to carry out longitudinal studies to investigate how students manage online groupwork over time, and how groupwork management is affected by various variables at the individual and group levels. In addition, although several studies on online groupwork implied that groupwork management in online environments had a positive influence on groupwork performance (Choi & Kang, 2010; Koh & Hill, 2009; Oliveira, Tinoca, & Pereira, 2011), there is a need to more explicitly link groupwork management to quality of groupwork in a longitudinal design.

There is also a need to conduct studies such as this with younger learners (e.g., secondary school and undergraduate learners), as findings from the present study and related findings on secondary homework (Xu & Wu, 2013) suggest that the relative influence of task importance, interest, and appeal on management strategies may be moderated by students’ age. Furthermore, it would be informative to examine how students manage online groupwork in a shared document platform, as the use of such a platform may influence how students approach online groupwork (e.g., feedback and help seeking). Finally, it would be valuable to open a new line of investigation to better understand and address the challenges associated with online group management in cross-cultural environments, because one’s attitudes toward online groupwork (e.g., perceived importance of doing online groupwork) may be influenced by cultural differences regarding the value of individual autonomy and choice (Eccles, 2005) and of the value structure of individualism and collectivism (Hofstede, 2001).

Conclusion

One research area increasingly gaining importance in the field of online collaborative learning is relating to students’ groupwork management (Choi & Kang, 2010; Rovai, 2007; Smith et al., 2011; Volet et al., 2009). The present investigation examined multilevel models of factors to predict online groupwork management, informed by the self-regulation literature and studies pertaining to online groupwork management.

Results revealed that, at the student level, online groupwork management was positively associated with previous online courses, feedback, peer- and learning-oriented reasons, and help seeking. In addition, older students more frequently took initiatives to manage online groupwork. Finally, groupwork management was positively related to help seeking and feedback at the group level.

In light of the lack of empirical research investigating a broad range of variables that contribute to groupwork management at the individual and group level, it can be argued that the present investigation expands and extends previous research on online collaborative learning (e.g., the importance of help seeking and feedback at both the student and group level). We hope that that our findings are of considerable importance and utility to academic researchers and online instructors who are interested in online collaborative learning, with online groupwork in particular.

References


From Learning Object to Learning Cell: A Resource Organization Model for Ubiquitous Learning

Shengquan Yu*, Xianmin Yang, Gang Cheng and Minjuan Wang

Institute of Modern Educational Technology, Beijing Normal University, China // Institute of Education, Jiangsu Normal University, China // College of Education, National Open University, China // Department of Learning Design and Technology, San Diego State University, USA // yusq@bnu.edu.cn // yangxianmin8888@163.com // chenggangouc@163.com // mwang@mail.sdsu.edu

*Corresponding author

(Submitted August 21, 2013; Revised December 9, 2013; Accepted February 6, 2014)

ABSTRACT

This paper presents a new model for organizing learning resources: Learning Cell. This model is open, evolving, cohesive, social, and context-aware. By introducing a time dimension into the organization of learning resources, Learning Cell supports the dynamic evolution of learning resources while they are being used. In addition, by introducing a semantic gene (knowledge ontology) into the model, Learning Cell can flexibly describe the internal structure and external relations of learning resources, allowing the evolution of learning resources to occur in an orderly way. Furthermore, by employing a computational model of social cognition network, Learning Cell enables not only materialized resource sharing but also the sharing of social cognition networks. Finally, by separately deploying resource structures and resource content in the cloud storage model, Learning Cell achieves context awareness of ubiquitous learning (u-learning) resources. Learning Cell represents a resource aggregation model that is different from the learning object model. It makes up for the defects of existing learning technologies in the following areas: the sharing of process information and social cognition networks, the intelligence of resources, and the evolution of content. Learning Cell provides a theoretical framework of u-learning resource organization.

Keywords

Ubiquitous learning, Learning Cell, Learning object, Knowledge relationship network, Aggregation model, Context awareness

Introduction

There have been three generations of resource organization models in the field of e-learning (See Figure 1): Integrable ware (Reusable Learning Unit), Learning Object, and Learning Activity. The Integrable-ware model (Li, 1997) emphasizes the combination and reuse of small learning resource units. A learning object is any digital resource that can be reused to support learning (Wiley, 2000). The main idea of learning objects is to break educational content down into small chunks that can be reused in various learning environments, in the spirit of object-oriented programming. The learning object model (ADL, 2004) provides solutions for data exchange between different learning management systems (LMS), allowing structured and interactive learning objects to be shared between LMS. The learning activity model (Britain, 2004) supports high-level sharing of learning processes and activities by reusing learning methods, learning strategies, and learning activities. The emergence of IMS-LD (IMS Learning Design) (IMS GLC, 2003) extended the sharing of learning resources from learning objects to learning activities, signaling a shift in learning-resource sharing from a technological problem to an educational problem.

However, the sharing of learning activity designs is not the ultimate goal of learning resource sharing. New learning technologies aim to cater to ubiquitous learning. Figure 1 also lists questions that need to be addressed in the future development of learning-resource sharing, for instance, how to share generative information during the learning process and how to adapt learning content for personalized learning.

U-Learning’s demands on learning resource organization

Pervasive computing creates a seamless computing space encompassing both the real world and the virtual world, where ubiquitous learning (u-learning) is well supported. Generally, u-learning allows people to access learning...
resources anytime anywhere. Context-aware u-learning is supported by mobile devices, wireless communications, and sensor technologies.

![Diagram showing the development of learning resource sharing](image_url)

**Figure 1.** The development of learning resource sharing

As one of the ideal models for future life-long learning, u-learning provides an environment in which learners can access information, communicate and engage in learning activities without temporal or spatial constraints. Learning in such environments has different characteristics than traditional learning (Hwang, Yang, Tsai, & Yang, 2009; Ogata & Yano, 2004; Wong & Looi, 2011; Yu, 2007): (1) u-learning may happen anytime and anywhere; (2) u-learning is personalized, context-aware and problem-oriented, and it provides suitable learning content and services to meet learners’ demands; (3) u-learning is embedded into daily life and may be decentralized and fragmented; (4) u-learning is socialized, with social interactions being an important component.

The core motivation behind u-learning is not the ubiquity of devices and connections, but the changes in learning modes brought about by pervasive computing. Traditional, statically hierarchical organizational structures cannot meet the new demands created by these changes. U-learning needs to provide the following new organization modes for learning resources:

- U-learning needs adaptive and contextual learning resources.
- U-learning requires a large number of learning resources, which are created and shared by the learners.
- U-learning needs evolving learning resources. Ubiquitous learning resources should exist in real time, reflect the latest developments in related areas and meet learners’ actual real-time needs. Such resources need to keep track of the information generated during the learning process. This information becomes the seed of resource evolution and reflects the history of knowledge construction.
- U-learning needs learning resources that are integrated with learning activities. Learning is not just acquiring information but is a process of effective internalization. This is achieved by providing learners with opportunities to participate in learning activities.
- Above all, u-learning needs the “human” resources that can be created by sharing social cognitive networks. Interaction in u-learning is not merely the interaction between learners and materialized resources, but also among the learners.

**Related work**

Existing learning resource standards, such as IEEE LOM (Learning Object Model) and SCORM (Sharable Content Object Reference Model) focus on describing and packaging resources, and their primary goal is to enable resource sharing and reuse across different platforms. Existing resource standards that are designed for traditional e-learning are characterized by knowledge transfer tools and are not suitable for new learning methods such as collaborative
learning and discovery learning, where resources can be collaboratively developed and shared using social networks. Therefore, some international research groups have initiated work on the next-generation model for sharing learning resources. For instance, the Instructional Management System Global Learning Consortium (IMS) has published the Common Cartridge V1.0 standard (IMS GLC, 2008). In addition, Advanced Distributed Learning (ADL) is working on a standard called SCORM version 2.0 (Allen, 2008).

Common cartridge

Newly proposed by IMS, IMS Common Cartridge is a resource package format specification integrated with IEEE LOM, Dublin Core, IMS Content Packaging, IMS QTI, and IMS Authorization Web Service v1.0. The design concept of Common Cartridge is to integrate learning resources of different types and from different sources to create a learning platform (Cheng, Xu, & Yu, 2009). Resource types include web pages, associated content, QTI tests, QTI question banks, discussion topics, and web links that can be stored in learning platforms or accessed anywhere on the Internet. Access control mechanisms are adopted to protect important content. Figure 2 describes the architecture of Common Cartridge.

![Figure 2. Architecture of Common Cartridge (IMS Global Learning Consortium, 2008)](image)

Compared with SCORM, IMS Common Cartridge represents progress because it expands the types of sharable resources to tests, allows discussions, and provides a mechanism for distributed resource integration. Common Cartridge’s package interchange file not only contains metadata descriptions and allows static resources to be organized and sorted but also supports a variety of resources such as external web pages, web links, discussion topics, project assessments, and question banks. In addition, Common Cartridge offers strategies for integrating external learning tools.

SCORM 2.0

SCORM is currently the most widely used learning resource package standard and has significant influence on both industries and academia. However, with the development of new technologies and concepts, many creative and prosperous Internet applications have been created in web 2.0 mode. Therefore, SCORM can no longer satisfy the various needs of e-learning and especially of u-learning. In response, ADL initiated work on SCORM 2.0, which is now overseen by LETSI (Learning, Education, and Training Systems Interoperability). The main goals of SCORM 2.0 are to support existing and emerging technologies, to support multiple teaching and learning modes, and to support multiple learning environments and modes, such as classroom learning, vocational training, distance education, self-learning, etc. (LETSI, 2008).
At present, different organizations and individuals have presented more than 100 white papers on SCORM 2.0. The content of these white papers covers the generic model, architecture, group learning, sequence and navigation, metadata, content integration, and assessments. Many papers suggest that SCORM should be simplified and call for the support of SOA architecture, Web 2.0, external resources and services, and user-defined features. The common suggestion (among 29 papers) is to include the support of SOA architecture and cross-platform integration. In addition, 13 papers emphasize the aggregation of external content, multi-format display, assessments, and 14 papers highlight the importance of learning designs.

Analyses

Both SCORM 2.0 and IMS Common Cartridge indicate that international standardization groups have made significant contributions to the development of learning technologies. However, both resource-sharing methods are not suitable for u-learning, because (1) they fail to plan and design tools that would enable the renewability of these resources, (2) they ignore human factors in resource organization (Yu & Yang, 2007), (3) they are unable to meet informal learners’ needs for individualized and diversified learning, and (4) the limited resource-sharing supported by SCORM, IMS learning design standards and Common Cartridge norms prevents resources from being imported and exported between different LMS. Real and accessible model should therefore set a good example for observers. Generative information produced during the learning process is an appropriate object for observation that can promote learners’ learning. Thus, this generative information should be included among the sharable learning resources.

An increasing number of researchers and practitioners realize that learning concepts and methods need to be revamped to represent the digital age. Connectivism, considered “the learning theory of the digital age” (Siemens, 2004), suggests that individuals are not able to possess all the knowledge they should because of the rapid change of knowledge bases and the increasing amount of available knowledge. Knowing how to obtain the knowledge we need and the ability to discern the importance of the information we find is now more important than acquiring static knowledge. Thus, learning encompasses not only the individual’s intrinsic psychological activities but also a process of linking to inside and outside nodes for storing knowledge and information. The current learning resource technologies focus only on the development and use of materialized resources while neglecting the possibilities of humanized resources—social interpersonal networks and cognitive networks.

Based on the above problems, this study attempts to go beyond learning objects and proposes a new sharable learning resource model for u-learning.

The proposed model: Learning cell

Because they are based on centralized storage and hierarchical directory structures, existing learning resource organizations cannot meet the demands of u-Learning. In this paper, we propose a new learning resource describing and packaging scheme called Learning Cell.

Principles of Learning Cell

Based on connectivism and collaborative knowledge construction theories, Learning Cell features cloud storage, a dynamic structure, and non-static data elements. Learning Cell is a learning resource that is open, generative, evolvable, connected, cohesive, intelligent, adaptive, and social (Yu, Yang, & Cheng, 2009). It is developed from the learning object model and designed for u-learning. The basic idea is to introduce a time dimension and an interpersonal cognition network into learning resources to make the learning resource evolvable. During the evolution process, generative information and the revision history are recorded, an interpersonal network is generated, and humans and knowledge connect with each other to form a knowledge network, which allows students to construct knowledge, understand the context of that knowledge, and share collective wisdom through social cognition networks.
Figure 3 shows how learning cell can support u-learning. To achieve pervasiveness and context awareness, the two important characteristics of u-learning, the learning space needs to be based on an educational cloud platform, which consists of pervasive network connections, content-aware devices, evolvable learning resources, cognition networks, learning designs, and community networks. The platform hosts authentic knowledge and field experts and connects these experts in a social wisdom network, establishing knowledge networking social services (KNS) using collective wisdom-based algorithms. As shown in Figure 3, a seamless learning space for u-learning can be achieved with the coordination of cloud-computing and multi-media technologies.

When users encounter problems or develop an interest in something, the intelligent terminals can detect their needs and communicate these needs to the educational cloud-computing platforms through pervasive communication networks. Based on users’ learning requests, the platforms would perform searches, computations, and transformations to choose the most appropriate learning content for the user and attach the content to learning services and knowledge networks. Learners would be able to contact other learners and experts who are interested in the same content to form learning communities, giving learners access to the most authoritative knowledge in the field and allowing learners to build relationships with experts. Such a learning model does not mimic traditional classroom learning in which there is one teacher and many learners, but is instead a 1:1 or even n:1 model of learning, in which many authoritative experts and collaborators serve as teachers for one learner.

**Structural model of Learning Cell**

Learning Cell uses the cloud storage model to provide resources for u-learning. Figure 4 shows the learning content and its structure. Learning Cell consists of components such as metadata, ontology, content, activities, assessments, generative information, and multi-formats. All these components connect with education cloud services through a variety of service interfaces (e.g., learning activities and assessments). In education cloud services, there is an immense amount of learning resources and records of learning, including activity records, editing records, evaluation records, usage records, learning communities, and other information generated during the learning process. The learning content itself can be stored in resource servers all over the world. When learners enter u-learning
environments, what they access is not unchangeable learning content but the dynamic structure of the learning content. When learning cells are accessed in different contexts, learning content is also aggregated differently to meet the learners’ needs.

Learning Cell is a learning service that can be accessed through a URL, which allows it to provide users with context-appropriate learning content and applications. Learning cells are goal oriented and can exist independently or be connected into a personalized knowledge network. The network contains metadata, aggregation models, knowledge ontology, learning content, learning assessments, learning activities, generative information, learning service interfaces and other resources, which are described in detail below.

Metadata is used to describe the attributes of Learning Cells so that they can be easily categorized, indexed, and shared.

Aggregation models prescribe the inner components and connection modes of Learning Cell. In contrast with Learning Object, Learning Cell adopts a semantic-based network aggregation model in which different components dynamically connect with each other to form a network. Different learning resources can be aggregated into one learning cell, and different learning cells can be aggregated into larger-scaled knowledge clouds. Knowledge ontology describes the basic concepts and inter-concept relationships of related knowledge, which are used to construct effective aggregation models and promote the aggregation of Learning Cell components and the dynamic connection of learning cells containing similar content.

Content is the main component of Learning Cell. Learners use content to obtain learning resources and to build their own knowledge. The content in a learning cell needs to have specific subjects and goals and to be independent and self-contained, even with small granularity. Assessment is used to examine how learners grasp learned knowledge and to adjust learning strategies based on the assessment results. Learning Cell records all of a learner’s interactive information and forms assessment reports after performing information analyses. Activities promote deep interactions between learners and content; they are process-oriented and enable the sharing of learning processes, strategies, and activities. Generative information is generated during the process of using Learning Cell. It contains user information, interaction information, Learning Cell revision history, and so on. Multiple formats indicate ways...
through which learners access a Learning Cell, including http, e-book, knowledge map, and video or audio. These formats allow a learning cell to be displayed on different devices.

Service interface is the primary channel of information exchange between Learning Cell and the cloud-computing environment. It defines the interface both to acquire and update process information when learners perform learning activities through a learning cell and to update the inner components and structure of learning cell. For example, when a learner revises a paragraph of the learning content and submits new related learning resources, this information is labeled and stored in a learning cell and can be shared with other learning systems.

The cloud storage architecture embodies the importance of distributed computing and cloud computing to learning technologies. Future learning resources would not be simply deployed from a centralized inner server but distributed and connected around the world. Learning resources on those nodes would be dynamically connected, based on rich semantics, forming a comprehensive, intelligent network.

Core features of Learning Cell

Learning Cell retains the accessibility, adaptability, affordability, durability, interoperability, and reusability of Learning Object but has, in addition, the following five unique properties (as shown in Figure 5):

Openness

Open-learning resources provide not only open access but also open content. Unlike the traditional static and closed-resource organization model, Learning Cell is built on a dynamic resource structure. Learning Cell allows resources to be updated with usage information, to grow by absorbing valuable online content such as Gadget, to interact with the external learning environment, and to be integrated with learning activity designs. Each learning cell is equipped with a learning system service interface, allowing the learning process to be traced and information to be exchanged between Learning Cell and the runtime environment.

![Figure 5. Open-structure model of Learning Cell](image)

Evolvable

Traditional learning content is static and difficult to update. Learning Cell makes learning content evolvable by updating content with the feedback it receives from the learners. Based on the idea of Web 2.0, learning content is no longer generated by a few professionals but is generated and updated by public users who are connected via a network. Unlike Learning Object, Learning Cell saves not only predefined course content, exercises, and activities
but also information generated during the learning process, such as submitted work, annotations on learning content, and learning records. Revisions to learning content are also recorded to reflect the evolving process of learners and learning cells.

Users first generate learning resources and invite collaborators to edit the open content. As the content evolves, more users familiarize themselves with the content and add comments and annotations to it. With time, as the users’ collective wisdom grows, the content will grow to meet learners’ needs. At present, the evolution of most open content is achieved by collaborative editing, and a version-control mechanism protects the content.

During the evolving process, learning resources are manually (users generate inter-resource relationships through an editor) or automatically (the system creates semantic relationship mining using semantic correlators) connected via semantic relationships, including similar relationships, hyponymy relationships, precursor relationships, inclusive relationships, and equivalence relationships. As the connections among resources strengthen, similar resources form subject resource groups, from which learners access the subject’s knowledge relationships. Meanwhile, users can organize semantically related resources into structured courses according to the inner-logic relationships of the knowledge. In the end, all resources become nodes in the resource network and are enhanced by forming dynamic connections with other resources. In addition to connections with materialized resources, human resources are also important. Interpersonal networks are formed when users perform various learning-related activities on the resource.
Learning Cell is cohesive, organizing all elements of the learning process into an orderly whole. The “gene” of a learning cell is formed according to an ontology-based knowledge structure and aggregation model, which controls the evolution and development of each learning cell. This structured learning resource organization is distinctive from loosely organized web resources such as blogs and BBSs. The network generated by Learning Cell is not a simple accumulation of learning resources, but a dynamic knowledge network.

The most important distinction between Learning Cell and Learning Object or SCORM-based online courses is the application of semantic network and ontology technology, which makes Learning Cell resemble an organism that grows and evolves under the control of an internal gene. Except when it exists as an independent and complete learning unit, each learning cell could serve as a node in resource networks and could connect with other nodes based on certain rules.

![Figure 8](image.png)

**Figure 8.** (a) Visualized editing of semantic relationships between learning cells; (b) semantically based visualized navigation

Because learning content is infused with the wisdom of all learners, the combination of physical resources and people would create a dynamically evolving and developing social cognitive network. Learners could acquire not only existing knowledge but also learning methods and knowledge acquisition channels.

Therefore, the social characteristics of future learning resources should be gradually strengthened. In u-learning, learning resources should serve as the bridge to connect learners in addition to communicating knowledge. The cognitive network attached to learning resources is an indispensable attribute of future learning resources. Aligning with this concept, learning cells include not only materialized learning resources but also interpersonal networks generated during the evolution process. Learning cells can be regarded as intermediaries between learning channels and social wisdom. Learning cells connect humans through learning content and generate a social cognition network in ubiquitous learning space.

Unlike the common conception of social networking, a social cognition network consists of knowledge, humans, and the interactions between them. In a social cognition network, a learning resource can be an independent learning unit connected to another resource via semantic relations. Through the social cognition network, learners can not only access materialized resources, but they can also find related people. For example, a learner can find subject-related experts and learning partners through learning content and access knowledge or promote learning activities from the experts or other learning partners.
The core feature of u-learning is its context awareness. It can adapt the learning services to the learning contexts; in other words, it can perceive users’ demands using intelligent learning devices and offer the most suitable learning modes and services. To realize such a learning model, we must improve the adaptability of learning devices and redesign the aggregation model to make learning resources more adaptable to these devices.

The context awareness of learning resources lies in the following two aspects: (1) intelligent adaption to learning terminals and 2) adaptability of learning content. The context awareness of Learning Cell arises from three features: 1) the predefined multi-formats (including different file types, resolutions, and code rates) of resources, through which Learning Cell can adapt learning resources to different learning terminals; (2) the addition of semantic information to resources, through which Learning Cell allows learning resources to be easily recognized and processed by intelligent devices so they can be indexed, matched, connected, and assigned; (3) the dynamic structure through which resources can be adjusted. To realize the dynamic aggregation and self-adaptation of learning content, we must change the static structure model so that all learning content is organized in a process-oriented, logical structure and is dynamically generative.

**Implementation of Learning Cell**

The success of Learning Object relies on SCORM-supported learning management system and IMS-LD-supported learning platform. In comparison, Learning Cell operates independently of a specific supporting environment.

**Runtime environment architecture (REA)**

Figure 9 shows the architecture of the Learning Cell runtime environment. Key components include the message transfer controller, resource locator, repository, learning cell runtime engine, active adaptor, and learning service interface. A u-network is a ubiquitous network that supports u-learning; it is accessible via the Internet, wireless communications, and digital TV networks, to which users can easily connect with a device. A u-network manages the data transfer and device communication that are necessary to transfer content among learning cells. Various display devices, the u-network, and the Learning Cell runtime environment work collaboratively, forming a Learning Cell-based, seamless learning environment.

The message transfer controller receives user request information from the ubiquitous network (u-network), analyzes the information, and decides where to send the information. The resource locator manages resource indexes, searches learning resources at users’ requests, and locates the resources in the repository. The repository stores learning cells and other resources, including generative information, semantic relationships, user information, and information about various devices. The Learning Cell runtime engine manages the information exchange between learning cells and the external environment; it consists of a series of APIs, including the Learning Cell grow API, the Learning Cell divide API, and the Learning Cell track API. Through this engine, learning cells content can evolve, resources can be aggregated based on semantics, and the learning cells themselves can divide and grow. The active adaptor receives device information from the message transfer controller; analyzes device types, screen sizes, and screen resolutions; and transforms the content into the most appropriate format for display on devices. The Learning Service interface provides users with a series of learning services, including learning tools, learning communities, learning content, learning activities, learning assessments, learning records, and semantic associations. Learners can use these interfaces in the u-network and access learning support services from any location and at any time.

The Learning Cell runtime environment is based on Java 2 Platform Enterprise Edition (J2EE) and Service Oriented Architecture (SOA) and can be divided into four layers: the repository layer, the service layer, the application layer, and the display layer.

The repository layer stores various data from the runtime environment and includes (1) a resource repository, which stores all of the resources, including learning cells and knowledge groups; (2) an ontology repository, which stores all of the knowledge ontology in the environment, including predefined ontology and user-generated ontology; (3) a
user-information repository, which stores information like user portfolios and trust degrees; (4) an activity repository, which stores information like discussions, voting patterns, and reflections; (5) a tool repository; and (6) a log repository, which stores logs from learning cells, knowledge groups, learning activities, and user operations. There are mainly two ways to generate knowledge ontologies. One is to import external ontologies applied widely in many organizations directly. The other is to build a collaborative ontology-creating environment for any user participating in producing different knowledge ontologies. In the latter, a strict mechanism should be provided to guarantee the quality of ontologies. After time testing, these ontologies could be exported to Web Ontology Language (OWL) file for public sharing in other organizations.

Figure 9. Architecture of the Learning Cell runtime environment

The service layer provides various services based on data from the repository layer, including learning activities, learning assessments, learning tools, version controlling, collaborative content editing, resource management, resource aggregation, resource indexing, format matching, and learning activities. During the process of collaborative content editing, we can not only adopt Wikipedia’s content audit mechanism, but also use semantic technology and trust evaluation technology (Yang, Qiu, Yu, & Hasan, 2014) to realize automatic content checking (Yang & Yu, 2013). Resource aggregation can be implemented by counting semantic similarities among resources and human selection and reorganization.

The application layer provides applications for users by calling services from the service layer. The applications include learning cells, knowledge groups, knowledge clouds, learning tools, personal space, and learning communities, all of which offer varied learning experiences.

The display layer automatically converts the format of learning cells according to the information provided by the display devices, which could include digital TVs, computers, smart phones, public information terminals, and live telecasts so that learning cells can be properly displayed on different devices. We should provide a variety of content format templates, so as to adaptively display the content on different devices.

Functions of the application layer

Figure 10 shows the functional model of the application layer. Its main functions are knowledge group (KG), knowledge cloud (KC), learning cell (LC), learning tool (LT), personal space (PS), and learning community (LC).
The Learning Cell function assembles all of the learning cells in the environment. Each learning cell is a resource entity; it can be a lesson or a knowledge point. A learning cell contains not only learning content but also learning activities, KNSs, semantic information, and generative information. Learning cells are available in multiple formats, such as web pages, e-books, concept graphs, and 3D models. Learning cells can introduce related assistant learning tools to support u-learning.

The knowledge group (KG) function assembles all of the knowledge groups in the environment (see Figure 12). Each knowledge group consists of learning cells on related subjects. For example, a course can be a knowledge group and each lesson or knowledge point in the course can be a learning cell. When users access the knowledge group, they can find all of the learning cells related to the course.

The knowledge cloud (KC) function aggregates multiple knowledge groups. Different knowledge groups are connected via semantic relationships. In a knowledge cloud, users can easily find all of the knowledge groups related to their subject.

The learning tool (LT) function assembles all of the personalized learning gadgets. In LT, users can not only preview or save gadgets but also upload gadgets. All gadgets conforming to open social standards can be integrated into LT. These gadgets can be used by learning cells, knowledge groups, personal space, and learning communities. For example, to enhance learning efficiency, some gadgets, such as translating gadgets, can be integrated into the learning content during the content creation or editing process.
Figure 12. A knowledge group consisting of multiple semantically related learning cells

The knowledge cloud (KC) function aggregates multiple knowledge groups. Different knowledge groups are connected via semantic relationships. In a knowledge cloud, users can easily find all of the knowledge groups related to their subject.

The learning tool (LT) function assembles all of the personalized learning gadgets. In LT, users can not only preview or save gadgets but also upload gadgets. All gadgets conforming to open social standards can be integrated into LT. These gadgets can be used by learning cells, knowledge groups, personal space, and learning communities. For example, to enhance learning efficiency, some gadgets, such as translating gadgets, can be integrated into the learning content during the content creation or editing process.

The learning community (LC) function assembles all of the learning communities in the environment. A learning cell is a collective learning environment (CLE) in which community members communicate, collaborate, or share with each other. Community members can publish a notice, initiate a discussion, share interesting resources, and initiate learning activities. Learning communities are related to LC, KG, and LT, and related learning cells, knowledge groups, and knowledge tools can be introduced into learning communities. In addition to learning communities, all users have their own personalized learning environment.

Personal space (PS) is the personal learning environment (PLE) of each user, containing functions for personal resource management, friend management, schedule management, gadget management, and personalized learning recommendations. In personal space, users can post basic personal information, manage (create, collaborate, and subscribe to) interesting learning cells and knowledge groups, and select recommended learning resources.

Applications and evaluations

As a proof of concept, our lab created the Learning Cell System (LCS), an open-knowledge community developed for u-learning. It supports collaborative knowledge editing, knowledge aggregation and evolution, multiple-level interaction, and multi-dimensional communication. Specifically, LCS allows the orderly evolution of resources, facilitates shared cognition networks and the collaborative construction of ontologies, and provides open-service tools. LCS can be accessed at http://lcell.bnu.edu.cn. Since its launch in May 2011, 8,851 users have registered, 15,793 learning cells have been created, 80 learning applets have been generated, and 1,507 knowledge groups and 116 learning communities have been formed (as of July 23, 2013).
Case study

We carried out two practical applications of LCS. The first application was designed to support e-learning for graduate students (formal learning), and the second one was to support regional collaborative teaching research for primary and secondary school teachers (i.e., informal learning). In the first e-learning case study, a course was offered to 25 graduate students through LCS (see Figure 13). The teacher created teaching content that allowed students to collaboratively edit them. The teacher designed all kinds of learning activities (discussion, test, vote, etc.) for students to participate. Students could use smart phones or tablet PCs to view learning content and interact, commenting, rating, and taking notes (See Fig. 14). In order to promote deep communication and knowledge sharing, the teacher created a course community. In this community, teachers could publish course information, put up some questions to guide students to in-depth thinking, assign individualized homework, and keep a watchful eye on students’ learning process. Students could upload their work, share course materials, and attend course discussions.

![Figure 13. E-learning process based on LCS](image)

![Figure 14. Students using tablet PCs (left) and a smart phone (right) to learn on LCS](image)

The case study was performed in Feixi County, Anhui Province, with 50 participating teachers in 10 schools. The collaboration process is shown in Figure 15. In the regional collaborative teaching research case study, teachers shared experiences through LCS. The director established several knowledge groups for the focus subjects and invited teachers to collaborate on the research. Meanwhile, learning communities were set up to encourage teachers to share their resources, knowledge, and experiences. Any teacher can create, edit, and comment on lesson plans anytime, anywhere with the mobile client APP of LCS (See Figure 16). Teachers could also publish their own teaching thoughts in a timely manner and communicate with others to improve together.
1. Platform training (including theory & technique)
   Determine the teaching and research tasks

2. Make teaching and research plan
   Divide interschool teachers into groups

3. Teachers construct knowledge independently
   Submit scheme

4. Collaboratively construct scheme within group
   Networked Communication
   Modify notation

5. Regional teaching and research class
   Implement teaching according to the scheme

6. Reflection after teaching
   Improve the scheme

7. Summarize the regional teaching and research activity
   Make the next teaching and research plan

Figure 15. Regional collaborative teaching research process based on LCS

Figure 16. Screenshots of mobile APP of LCS

Evaluations

In the above two case studies, investigations were carried out to determine the LCS’s usability and user attitudes. The usability investigation was based on an SUS tool developed by John Brooke (Brooke, 1996), and the questionnaires were published using a professional investigation platform (http://www.sojump.com/). Fifty users participated in the investigation; the results are shown in Figure 17.

The results show that 68% of the users felt confident using LCS, 26% of the users were neutral, and 6% of the users were not willing to use the system due to their lack of confidence with it. In general, most users had positive attitudes toward LCS. Further investigations reveal that the non-confident users felt that LCS was too complicated.
The first group of user attitude investigations was carried out with the teachers involved in the collaborative research case study. Twenty-five questionnaires were sent out by email and 23 (92%) were returned. The results are shown in Figure 18.

![Figure 17. LCS usability investigation results](image)

**I felt very confident using the system**

<table>
<thead>
<tr>
<th>Feeling</th>
<th>Number</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly agree</td>
<td>34.00%</td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>26%</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>Disagree</td>
<td>34.00%</td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>2%</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 18. Investigation results of LCS-supported collaborative teaching research**

Figure 18 shows that 82.61% of the teachers liked LCS-supported collaborative research, and 8.7% of teachers disliked it. Further investigations indicated that some teachers felt LCS was too complicated to use without additional usage guidance.

The second group of user attitude investigations was carried out with the graduate students who participated in the e-learning course. Twenty-five questionnaires were distributed, and all of them were returned. The results of the general feedback are shown in Table 1, and the results of feedback about specific LCS functions are shown in Table 2.

<table>
<thead>
<tr>
<th>Options</th>
<th>Number</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functions are rich and practical</td>
<td>17</td>
<td>68%</td>
</tr>
<tr>
<td>Helpful for collaborative learning</td>
<td>20</td>
<td>80%</td>
</tr>
<tr>
<td>Resource organization is innovative and practical</td>
<td>19</td>
<td>76%</td>
</tr>
<tr>
<td>LCS compensates for the deficiencies of existing learning platforms and conforms to the development direction of learning technologies</td>
<td>17</td>
<td>68%</td>
</tr>
</tbody>
</table>

**Table 1. User feedback about LCS-supported e-learning (25 users)**

Table 1 shows that 68% of students approved the system functions and technologies, 80% of students liked LCS-supported collaborative learning, and 76% of students believed LC was innovative and practical. Overall, most students have a positive attitude towards LCS-supported learning.
Table 2. User feedback on LCS functions (25 users)

<table>
<thead>
<tr>
<th>Options</th>
<th>Number</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content is open to generative information, such as comments and</td>
<td>23</td>
<td>92%</td>
</tr>
<tr>
<td>annotations, and helpful for KNS construction.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learners can obtain the updated information on demand.</td>
<td>13</td>
<td>52%</td>
</tr>
<tr>
<td>Learning Cell is evolvable, dynamic, and open.</td>
<td>21</td>
<td>84%</td>
</tr>
<tr>
<td>Learning Cell provides multiple formats to choose from.</td>
<td>22</td>
<td>88%</td>
</tr>
<tr>
<td>Content, activities, and tools are integrated.</td>
<td>20</td>
<td>80%</td>
</tr>
<tr>
<td>Participating in collaborative editing helps me absorb others’ wisdom.</td>
<td>23</td>
<td>92%</td>
</tr>
<tr>
<td>KNS helps me find knowledge-related learners and experts.</td>
<td>19</td>
<td>76%</td>
</tr>
<tr>
<td>KNS promotes learning by helping me communicate with related learners</td>
<td>16</td>
<td>64%</td>
</tr>
<tr>
<td>and experts.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic-based resource organizations (knowledge groups) are flexible</td>
<td>19</td>
<td>76%</td>
</tr>
<tr>
<td>and help me find resources on demand.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge semantic networks within knowledge groups allow users to</td>
<td>18</td>
<td>72%</td>
</tr>
<tr>
<td>contribute to strengthening the relationships among different knowledge points.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows that 92% of students liked the features of openness and collaborative editing. More than 80% of students liked the features of evolution, multiple formats, and integration. More than 70% of students enjoyed using the features of knowledge semantic network, semantic-based resource organizations, and KNS. In addition, half of students were able to obtain the updated information on demand, and 64% of students believed KNS can promote learning.

Valuable user comments about the system include: (1) the system is somewhat complicated, and user experience can be improved; (2) the adaptability of LC to multiple devices is advantageous, but more formats should be supported; (3) Learning Cell updates can be synchronized with micro blogging.

The above investigations reveal that LCS is functioning but there are some problems. In the future, we will simplify LCS operations and perform more and deeper investigations on LCS application.

Conclusion

Effective learning requires an intelligent, seamless learning space. Compared with other learning environments, seamless learning space accommodates high immersion and mobility. The construction of such a learning environment cannot be achieved without communication networks; context-aware, intelligent terminals; learning resources; and education cloud centers. Without the proper organization of learning resources, the appearance of new technological environments will not create ideal learning effects. The organization of learning resources is a fundamental factor of seamless learning environments. Based on Learning Object, this paper proposes a new learning resource organization model, Learning Cell, which provides a theoretical and practical basis for resource organization in future u-learning. Learning Cell is open, generative, evolvable, connected, cohesive, intelligent, adaptive, and social.

Our research team has successfully developed the Learning Cell System (LCS), a runtime environment for Learning Cell and used the basic ideas of Learning Cell to support learning. LCS can provide technical support for the following four aspects of the application: (1) In school context, teachers and students can create an online course together. The system can save all the learning procedural information, so as to generate standard courses in conformity with the SCORM standard. Teachers can also collaborate to build the school-based resource center and share teaching experience with each other. (2) Users can achieve personal knowledge management and organizational knowledge management in PS. (3) When used in industries, employees can be attracted to actively participate in learning by designing rich activities. They can also use the learning community to aggregate network and wisdom to promote collaboration and communication. (4) In higher educational institutions, LCS can support the smooth implementation of blended learning.
Pilot testing reveal the initial success of the Learning Cell Model and the Learning Cell System. In the future, we will carry out further research on evolution control, contextual cognition, personalized resource recommendation, application modes, and ecology-based knowledge construction. We believe that Learning Cell can greatly contribute to the resource design and sharing mechanisms of u-learning’s development.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (project number: 61073100), the National Social Science Foundation of China for the Youth Scholars of Education(project number: CCA130134), and Jiangsu Qing Lan Project.

References


Analyzing Log Files to Predict Students’ Problem Solving Performance in a Computer-Based Physics Tutor

Young-Jin Lee
Educational Technology Program, University of Kansas, 1122 West Campus Road, Room 413, Lawrence KS, 66045, USA // yjlee@ku.edu

ABSTRACT

This study investigates whether information saved in the log files of a computer-based tutor can be used to predict the problem solving performance of students. The log files of a computer-based physics tutoring environment called Andes Physics Tutor was analyzed to build a logistic regression model that predicted success and failure of students’ problem solving from their past interactions with the computer-based tutor. The logistic regression model developed in this study was able to correctly identify about 70% of the observed problem solving performance. The 10-fold cross-validation and the Receiver Operating Characteristic (ROC) curve analyses suggest that the developed logistic regression model can predict students’ problem solving performance on unseen new problems with a similar accuracy in the future.

Keywords
Log file analysis, Computer-based learning environment, Problem solving, Statistical modeling

Introduction

With the rapid development of computer and telecommunication technologies, computer-based learning environments, such as Massive Open Online Courses (MOOCs), are gaining popularity especially in higher education (Allen & Seaman, 2010). One of the benefits of computer-based learning environments is that they provide instructors with a new avenue for understanding the learning processes of their students; since students leave hidden trails in the log files recorded in the computer-based learning environment, instructors can quietly collect information relevant to the learning activities of their students (e.g., how much time students spend on a particular learning content, when they work during the day/week, etc.) without interrupting their learning processes.

As computer-based learning environments providing log files become more prevalent, Educational Data Mining (EDM) is emerging as a new, exciting research field. One of the active research areas in EDM is estimating academic performance of students who are using computer-based learning contents. For example, Hwang and Wang (2004) reported that the academic performance of college students taking a technology course was positively correlated with diligence and intensity of their learning in an e-learning environment. Lykourentzou, Giannoukos, Nikolopoulos, Mpardis, and Loumos (2009) proposed a data mining technique that can identify at-risk students who are likely to fail or drop an online course. They used a variety of data mining techniques, such as Feed-Forward Neural Networks, Support Vector Machines (SVM) and probabilistic ensemble simplified fuzzy ARTMAP, to build their prediction model from log files of an e-learning environment. Palazzo, Lee, Warnakulassoriya, and Pritchard (2010) developed a computational method that can detect students who submitted copied answers to their electronic homework problems by analyzing the log files saved in a Web-based tutoring system. They found that college freshmen who copied more than 30% of electronic homework problems scored up to 2 standard deviations lower in the final exam than their non-copier classmates who had comparable academic skills measured at the beginning of the semester (Palazzo, Lee, Warnakulassoriya, & Pritchard, 2010). Macfadyen and Dawson (2010) were able to build a logistic regression model that correctly identified 81% of students who received a failing grade by examining log files of their learning management system that documented learning activities of their students, such as the number of email messages sent, discussion messages posted and assessment modules completed by students.

This study seeks to extend the body of knowledge in EDM by focusing on more fine-grained learning activities occurring in a computerized learning environment, and how such information can be used to predict the learning outcomes of students for a longer period of time. While previous studies focused on whether or not students visited a particular Web page, this study examines more specific learning activities taking place in a computer-based learning environment (e.g., whether students were able to solve relevant problems with or without using instructional supports available in the computer-based learning environment). Moreover, while previous studies focused on predicting one academic performance measured at the end of the semester (e.g., final exam score or course failure), this study tries to predict students’ academic performance throughout the semester (e.g., problem solving performance in weekly
The objective of this study is to build a statistical model that can estimate the probability for students to successfully accomplish a problem solving step required to resolve a difficult physics problem without using hints available in the computer-based tutor, given their interaction history recorded in the log files of the learning environment (e.g., whether students were able to solve related problems, when they solved those problems, how well they did and how much time they spent in solving those problems, etc.).

**Method**

**Andes physics tutor**

In this study, the log files from Andes Physics Tutor (http://andestutor.org) were analyzed to build a statistical model that predicted how likely a student is to correctly solve a physics problem from his or her past interactions with a computer-based tutor. Andes Physics Tutor is an intelligent tutor program capable of providing appropriate instructional supports and guidance while students are solving difficult physics problems requiring an analytic solution. Unlike many computer-based instructional programs, which typically provide a single corrective feedback or hint on the final answer even though students have to go through multiple steps before arriving at their final answer, Andes Physics Tutor provides appropriate feedback and allows students to ask for hints on each problem solving step. This instructional approach has been found to be very effective in helping students learn difficult physics concepts and improve their problem solving skills (VanLehn et al., 2005).

Figure 1 shows an example of physics problems provided in Andes Physics Tutor. While solving a physics problem in Andes Physics Tutor, students need to define physical quantities and variables, such as mass, velocity, acceleration, external force, etc., and choose an appropriate coordinate system for the given problem as if they were solving it on paper. Also, students have to specify the steps they find necessary while solving a problem. Andes Physics Tutor watches students, just like a human tutor, and provides appropriate instructional supports and guidance when students make a mistake during their problem solving processes (e.g., incorrectly define a physical quantity, choose a step that is not required to solve the problem or fail to resolve any required step) or in response to hint requests from students.
Data set

The data set analyzed in this study was obtained from the Pittsburgh Science of Learning Center (PSLC) (http://www.learnlab.org) which provides log files created from several computer-based learning environments they had developed in the past. Their DataShop Web service allows education researchers around the world to access log files depicting actions of students learning various subject matters, from foreign language to algebra and physics, in computer-based learning environments (Koedinger, Baker, Cunningham, Skogsholm, Leber, & Stamper, 2010). The data set analyzed in this study documents learning activities of 69 students enrolled in an introductory physics course at the United State Naval Academy (USNA) in Fall 2008 semester. The original data set obtained from the PSLC DataShop includes 175,267 database transactions where each transaction contains information about students, their learning activities and other information internally used in Andes Physics Tutor (Koedinger et al., 2010). Among these, seven variables relevant to the problems students solved and their problem solving processes were used to create explanatory variables that were employed in the statistical model developed in this study (see Table 1).

Table 1. Seven variables selected from the transaction record in the PSLC data set

<table>
<thead>
<tr>
<th>Variables from PSLC data set</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anonymized student ID</td>
<td>Anonymous student ID generated by PSLC’s DataShop Web service</td>
</tr>
<tr>
<td>Time</td>
<td>Time at which the transaction occurred</td>
</tr>
<tr>
<td>Student response type</td>
<td>The type of attempt made by the student (e.g., ATTEMPT or HINT_REQUEST)</td>
</tr>
<tr>
<td>Problem name</td>
<td>The name of the problem associated with the transaction</td>
</tr>
<tr>
<td>Step name</td>
<td>The name of a problem solving step associated with the transaction</td>
</tr>
<tr>
<td>Outcome</td>
<td>The tutor’s evaluation of the student’s attempt (e.g., CORRECT, INCORRECT or HINT)</td>
</tr>
<tr>
<td>Knowledge Component (KC)</td>
<td>Knowledge Component associated with the transaction</td>
</tr>
</tbody>
</table>

Knowledge Component (KC) is defined as “an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks” (Koedinger, Corbett, & Perfetti, 2010), which can be generalized to various cognitive constructs such as production rule (Newell, 1990; Anderson & Lebiere, 1998), schema (van Merriënboer & Sweller, 2005; Gick & Holyoak, 1983) or facet (Minstrell, 2001), and everyday terms such as concept, principle, fact or skill. In order to solve the physics problem shown in Figure 1, for example, students need to find out the acceleration of a car, which in turn requires an understanding of a few prerequisite physics concepts and problem solving skills such as (1) gravity and centripetal force, (2) setting up an appropriate coordinate system suitable for the problem, and (3) gravity as a centripetal force. These physics concepts and problem solving skills are the KCs required for calculating the acceleration of a car in the physics problem shown in Figure 1. PSLC researchers employed a statistical analysis technique called Learning Factor Analysis (LFA) to identify KCs for each problem solving step in all Andes Physics Tutor problems in the data set (Cen, Koedinger, & Junker, 2006).

Data pre-processing

Since each step in the Andes Physics Tutor problem is associated with one or more KCs, this study focused on students’ interactions with the computer tutor from the problems that provided an opportunity to learn relevant KCs in the past, rather than looking at all past interactions saved in the log file. For the problem shown in Figure 1, for instance, which has three KCs (a centripetal force, how to set up an appropriate coordinate system, and the relationship between gravitational and centripetal forces), this study examined students’ interactions with the computer tutor while they were trying to solve other problems requiring an understanding of at least one of these three KCs, instead of examining all of their past interactions recorded in the Andes Physics Tutor log file. In particular, this study looked at (1) how many times students had a chance to learn the required KCs in the past; (2) how long they spent before submitting their first answer; (3) whether their first attempt to solve relevant problems was correct or incorrect; (4) how many hints they requested; (5) how many correct answers they submitted and (6) how many mistakes they made while solving the problems requiring an understanding of at least one relevant KCs in the past.

Since all this information, except for the correctness of the submitted answer, is not directly available in the original data set obtained from the PSLC DataShop Web service (see Table 1), data pre-processing had to be performed. In order to facilitate data pre-processing, the downloaded data set, which was in a tab-delimited text format, was first
imported into a MySQL database (http://www.mysql.com), and a data pre-processing program capable of accessing MySQL database tables was written in Python programming language (http://python.org). The data pre-processing program executed a series of SQL (Structured Query Language) queries against MySQL database tables to select students’ past interactions with the computer tutor while trying to solve physics problems providing relevant learning opportunities. From 175, 267 transactions in the original log file obtained from the PSLC DataShop, the pre-processing program yielded a data set consisting of 35,549 records that were eventually analyzed in this study.

**Building a statistical model predicting correct first try of students**

Equation 1 shows a logistic regression model predicting problem solving performance of students, given their past interactions with Andes Physics Tutor.

\[
Pr_{\text{correct first try}}(y_i = 1) = \logit(y_i) = \frac{e^{\beta_0 + \beta_1 C + \beta_2 T_{\text{correct}} + \beta_3 T_{\text{wrong}} + \beta_4 N_{\text{hint}} + \beta_5 N_{\text{correct}} + \beta_6 N_{\text{wrong}} + \beta_7 N_{\text{hint}} C} + \beta_8 N_{\text{hint}} T_{\text{correct}} + \beta_9 N_{\text{hint}} T_{\text{wrong}} + \beta_{10} N_{\text{hint}} N_{\text{correct}} + \beta_{11} N_{\text{hint}} N_{\text{wrong}})}{1 + e^{\beta_0 + \beta_1 C + \beta_2 T_{\text{correct}} + \beta_3 T_{\text{wrong}} + \beta_4 N_{\text{hint}} + \beta_5 N_{\text{correct}} + \beta_6 N_{\text{wrong}} + \beta_7 N_{\text{hint}} C} + \beta_8 N_{\text{hint}} T_{\text{correct}} + \beta_9 N_{\text{hint}} T_{\text{wrong}} + \beta_{10} N_{\text{hint}} N_{\text{correct}} + \beta_{11} N_{\text{hint}} N_{\text{wrong}}}
\]  

(1)

Where, \( \logit(y_i) = \frac{e^{\beta_0 + \beta_1 C + \beta_2 T_{\text{correct}} + \beta_3 T_{\text{wrong}} + \beta_4 N_{\text{hint}} + \beta_5 N_{\text{correct}} + \beta_6 N_{\text{wrong}} + \beta_7 N_{\text{hint}} C} + \beta_8 N_{\text{hint}} T_{\text{correct}} + \beta_9 N_{\text{hint}} T_{\text{wrong}} + \beta_{10} N_{\text{hint}} N_{\text{correct}} + \beta_{11} N_{\text{hint}} N_{\text{wrong}}}{1 + e^{\beta_0 + \beta_1 C + \beta_2 T_{\text{correct}} + \beta_3 T_{\text{wrong}} + \beta_4 N_{\text{hint}} + \beta_5 N_{\text{correct}} + \beta_6 N_{\text{wrong}} + \beta_7 N_{\text{hint}} C} + \beta_8 N_{\text{hint}} T_{\text{correct}} + \beta_9 N_{\text{hint}} T_{\text{wrong}} + \beta_{10} N_{\text{hint}} N_{\text{correct}} + \beta_{11} N_{\text{hint}} N_{\text{wrong}}} \)

\( \beta \): Regression coefficients,
\( C \): Number of opportunities to learn KC,
\( T_{\text{correct}} \): Average correct duration time,
\( T_{\text{wrong}} \): Average error duration time,
\( N_{\text{hint}} \): Average hint requests,
\( N_{\text{correct}} \): Average correct answers,
\( N_{\text{wrong}} \): Average wrong answers submitted while solving problems requiring an understanding of relevant KCs in the past

The outcome variable, \( y_i \), is a binary variable indicating whether students were able to successfully accomplish a certain problem solving step in a particular Andes Physics Tutor problem at their first attempt without using hints available in the tutoring system. The statistical model shown above estimates \( Pr_{\text{correct first try}} \) based on 11 explanatory variables consisting of 6 main effects and 5 interactions. The main effect explanatory variables try to capture important information that can affect the problem solving performance of students. For example, \( T_{\text{correct}} \), the average correct duration time, measures the amount of time students spent before submitting their first correct answer while solving relevant problems in the past. Similarly, \( T_{\text{wrong}} \), the average error duration time, measures the amount of time students spent before submitting their first incorrect answer or requesting hints while working on the problems requiring an understanding of relevant KC in the past. All explanatory variables were log-transformed because they were severely skewed, as shown in Figure 2, and their range was varied quite significantly. The log-transformation made explanatory variables follow an approximately normal distribution and have a comparable range (see Figure 2 and Table 2). Also, in order to help interpret the regression coefficients of the fitted model, all explanatory variables were centered by subtracting the mean of the data, which made each main effect coefficient correspond to a predictive difference with the other explanatory variables held constant at their average values (Gelman & Hill, 2007).

The development of the statistical model followed the general guidelines suggested by Gelman and Hill (2007). First, the data were fit to the logistic regression model containing only the main effect variables that might be expected to be important, for substantive reasons, in predicting the response variable. Then, for the main effect variable that has a large effect, their interactions with other main effect variables were included in the statistical model to see if they can explain the variance that was not accounted for in the main effect model. Among 6 main effect explanatory variables, \( N_{\text{hint}} \), the number of hint requests made by students while solving relevant problems in the past, was found to explain the largest amount of variance in the data. Hence, the interaction terms involving \( N_{\text{hint}} \) were included in the final statistical model to examine whether such interaction terms can improve the predictive power of the logistic regression model.
Results

Descriptive statistics

Table 2 summarizes the descriptive statistics of the main effect variables in their linear and log-transformed forms. Students, on average, had about 10 opportunities to learn KCs for the current problem solving step in question. It is not surprising to find that students took more time before submitting their first answer when their answer was incorrect, compared to when their answer was turned out to be correct. Also, students requested about 3 hints, on average, while resolving a problem solving step requiring an understanding of relevant KC. As briefly mentioned above, the variables in their linear form are severely skewed. Therefore, Table 2 reports median and semi-interquartile range values for the original variables, instead of conventional mean and standard deviations.

Table 2. Descriptive statistics of main effect variables

<table>
<thead>
<tr>
<th>Main effect variable</th>
<th>Original</th>
<th>Log-transformed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Semi-interquartile range</td>
</tr>
<tr>
<td>$C_{KC}$</td>
<td>10.00</td>
<td>12.50</td>
</tr>
<tr>
<td>$T_{correct}$</td>
<td>11.88</td>
<td>7.66</td>
</tr>
<tr>
<td>$T_{wrong}$</td>
<td>38.10</td>
<td>27.19</td>
</tr>
<tr>
<td>$N_{hint}$</td>
<td>3.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$N_{correct}$</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$N_{wrong}$</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Logistic regression analysis

In order to investigate whether students’ past problem solving performance and interactions with the computerized tutoring system recorded in the log files can predict $Pr_{correct\ first\ try}$, the probability for students to get each problem solving step correct at their first attempt without using hints, a logistic regression analysis was conducted. Table 3 summarizes the parameter estimates and the standard errors for the final logistic regression model developed in the study. The regression coefficients for all main effects and interactions were statistically significant at least at the 1% confidence level, indicating that the explanatory variables the data pre-processing program had produced were indeed useful in estimating the likelihood of the success of students in resolving the current problem solving step in question.
Table 3. Logistic regression model: Parameters and standard errors

<table>
<thead>
<tr>
<th>Main effects</th>
<th>Parameter estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of opportunities to learn KC ($\beta_1$)</td>
<td>-0.20***</td>
<td>0.03</td>
</tr>
<tr>
<td>Correct duration time ($\beta_2$)</td>
<td>0.35***</td>
<td>0.01</td>
</tr>
<tr>
<td>Error duration time ($\beta_3$)</td>
<td>0.23***</td>
<td>0.03</td>
</tr>
<tr>
<td>Number of hint requests ($\beta_4$)</td>
<td>-0.75***</td>
<td>0.03</td>
</tr>
<tr>
<td>Number of correct answers ($\beta_5$)</td>
<td>0.15**</td>
<td>0.05</td>
</tr>
<tr>
<td>Number of wrong answers ($\beta_6$)</td>
<td>-0.11***</td>
<td>0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interactions</th>
<th>Parameter estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hint requests : Number of opportunities to learn KC ($\beta_7$)</td>
<td>-0.20***</td>
<td>0.01</td>
</tr>
<tr>
<td>Number of hint requests : Correct duration time ($\beta_8$)</td>
<td>0.23***</td>
<td>0.02</td>
</tr>
<tr>
<td>Number of hint requests : Error duration time ($\beta_9$)</td>
<td>0.21***</td>
<td>0.02</td>
</tr>
<tr>
<td>Number of hint requests : Number of correct answers ($\beta_{10}$)</td>
<td>-0.13**</td>
<td>0.05</td>
</tr>
<tr>
<td>Number of hint requests : Number of wrong answers ($\beta_{11}$)</td>
<td>0.32***</td>
<td>0.03</td>
</tr>
</tbody>
</table>

*p < .01.  **p < .001.

Evaluation of logistic regression model

In order to evaluate the discriminatory power of the developed logistic regression model, Percentage of Correct Classification (PCC) and c-statistic were computed. With the cut-off probability of 0.5, the developed logistic regression model was able to correctly identify 70.46% of the cases in the data set analyzed in this study. The model resulted in false positives (FP), classifying a student who in fact successfully resolved the problem solving step as ‘failed,’ at a rate of 13.60% (4,834 out of 35,549 cases) as shown in Table 4. The logistic regression model also resulted in false negatives (FN) 15.94% of the time by erroneously classifying 5,667 cases into the ‘successful’ group even though these students were in fact failed to resolve the problem solving step in question (see Table 4).

Table 4. Percentage of correct classification

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Resolved the step</th>
<th>Failed to resolve the step</th>
<th>Percentage correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolved the step</td>
<td>9,994</td>
<td>4,834 (FP)</td>
<td>67.40</td>
<td></td>
</tr>
<tr>
<td>Failed to resolve the step</td>
<td>5,667 (FN)</td>
<td>15,054</td>
<td>72.65</td>
<td></td>
</tr>
<tr>
<td>Overall percentage</td>
<td></td>
<td></td>
<td>70.46</td>
<td></td>
</tr>
</tbody>
</table>

Data mining researchers typically perform a stratified 10-fold cross-validation analysis in order to estimate an overall error rate of a statistical learning model. In this evaluation approach, the data set is randomly divided into 10 parts in which the positive and the negative classes are represented in approximately the same proportions as in the full data set. Each part is held out in turn and the statistical learning model is built on the remaining nine-tenths; then its error rate is calculated on the holdout set. Therefore, the model building procedure is executed 10 times on different data sets and the 10 error estimates are then averaged to yield an overall error rate (Witten, Frank, & Hall, 2011). In this study, the 10-fold cross-validation analysis produced an overall prediction rate of 0.704, which is close to the PCC value obtained from the full data set, suggesting that the developed logistic regression model was fairly robust and it can have a similar predictive power on unseen data sets in the future.

Although PCC is commonly used to measure the discriminatory power of a statistical model, its error rate depends on a single cut-off probability (Kleinbaum & Klein, 2010). To address this issue, people often conduct a Receiver Operating Characteristic (ROC) curve analysis which produces a summary statistic based on a range of cut-off probabilities, not just one cut-off probability. When applied to a logistic regression model, an ROC curve is a plot of false positive vs. true positive derived from the cut-off probabilities for predicted values. Since a good logistic regression model will report small false positive and large true positive rates, the area under an ROC curve, which is often called AUC or c-statistic, will become larger as its discriminatory power grows (Kleinbaum & Klein, 2010). The AUC/c-statistic can vary from 0.5 (discriminating power not better than simple guessing) to 1.0 (perfect discriminating power) and is known to be equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (Fawcett, 2006). In this study, the AUC/c-statistic was found to be 0.78, indicating that the logistic regression model developed in this study has a fairly good predictive
power for the data set used in the model building stage. In Figure 3, the black line is a ROC curve obtained from the logistic regression model of this study while the red dotted line represents a ROC curve with no discriminatory power (ACU/c-statistic = 0.5).

![Figure 3. The ROC curve obtained from the developed logistic regression model](image)

Since research on predicting students’ problem solving performance from information captured in the log files of a computer-based learning environment is in its infancy, only a handful of studies are available that allows us to compare the predictive power of the logistic regression model reported in this study. Pardos and Heffernan (2011) incorporated problem difficulties into a standard Bayesian Knowledge Tracing (BKT) approach (Corbett & Anderson, 1995) to yield ROC values of 0.67 for ASSISTment and 0.53 for Cognitive Tutor data. Pardos, Gowda, Baker and Heffernan (2011) were able to obtain a ROC value of 0.77 when they applied a neural network that has varying numbers of hidden layers. Most recently, Pardos, Bergner, Seaton and Pritchard (2013) got ROC values of 0.65, 0.53, and 0.54 when they applied modified BKT to homework, lecture problems and exams in an edX MOOC data. Obviously, more in-depth studies are needed to systematically compare the predictive power of a logistic regression to other data mining algorithms. But, these results seem to indicate that the logistic regression model developed in this study has a comparable or better predictive power than other approaches.

**Discussion**

**Main effects**

The regression coefficients for the main effect explanatory variables suggest that the time students spend before submitting their first answer to the problems requiring an understanding of relevant KC, regardless of the correctness of their answers, is positively correlated with better problem solving performance in the future. Adding 1 to the log-transformed average correct duration time, $T_{correct}$, when the other explanatory variables are held constant at their mean values, corresponds to a positive difference in the probability of getting the current problem solving step correct about 8.8%. Similarly, a unit increase in the log-transformed average error duration time, $T_{wrong}$, is associated with an increase of about 5.8% in the correct first try probability. As is expected, students who submitted more correct or fewer incorrect answers, and requested fewer hints while solving relevant problems in the past showed much better performance on the current problem solving step in question. In particular, the number of hint requests, $N_{hint}$, showed the largest predictive power in discriminating the success of students; adding 1 to the log-transformed
\( N_{hint} \), when the other explanatory variables are held constant at their mean values, would yield about 18.8% decrease in the probability of getting the current problem solving step correct (see Table 3).

Interestingly, the coefficient for the number of opportunities to learn required KCs, \( C_{KC} \), was found to be negative, indicating that the more chances students had to learn relevant KCs, the more likely they are to get the current problem solving step incorrect; the increase of 1 in the log-transformed \( C_{KC} \) decreases the probability for students to get the current problem solving step correct by about 5% when the other explanatory variables are held constant at their mean values (see Table 3). At first glance, this result looks counter-intuitive because it would be reasonable to assume that students should be able to extract a schema while solving relevant problems in the past, which would in turn help them resolve the current problem solving step at hand. One possible explanation for this seemingly counter-intuitive result is that students with higher \( C_{KC} \) might be academically weaker students. Hence, these students did not do well even though they worked on more Andes Physics Tutor problems in the past. This interpretation is in part supported by the fact that the USNA students were allowed to solve as many Andes Physics Tutor problems as they found it necessary; the number of Andes Physics Tutor problems the USNA students solved in this semester varied greatly, from 1 to 74, indicating that Andes Physics Tutor was used as an add-on to regular learning activities, such as lectures and written problem sets. As briefly mentioned earlier, Andes Physics Tutor provides appropriate feedback and instructional supports while students are trying to solve difficult physics problems, which would make it more appealing to academically weaker students who need extra help. Further data analyses (e.g., examining academic performance of students measured independently) would be needed in order to find a more definite answer to this question.

**Interaction effects**

The negative coefficient for the interaction between the number of opportunities to acquire relevant KCs, \( C_{KC} \), and the number of hint requests, \( N_{hint} \), indicates that the importance of the number of opportunities to learn relevant KCs as an explanatory variable increases for students who used more hints while solving relevant problems in the past (and vice versa); for a unit increase in the log-transformed number of hint requests, the additional value of \(-0.20\) is added to the coefficient for the number of learning opportunities students had, which is also negative. As a result, the probability of submitting a correct answer for the current problem solving step decreases slightly faster when students used more hints while solving relevant problems in the past.

Also, the positive coefficients for the interactions between the number of hint requests, \( N_{hint} \), and the duration times, \( T_{correct} \), \( T_{wrong} \), suggest that the negative main effect of the number of hint requests would be mitigated if students spent more time in solving the relevant problems before submitting their first answer, regardless of their correctness. In other words, the logistic regression model predicts that the students who spent more time before submitting their first answer, among students who requested the same number of hints, are more likely to successfully resolve the current problem solving step at their first attempt without requesting hints. Previous study has found that learners often abuse instructional supports available in the compute-based learning environment without actually engaging in learning (Baker, Walonoski, Heffernan, Roll, Corbett, & Koedinger, 2008). The positive coefficients for the interactions between the number of hints and the duration times seem to support Baker et al. (2008)'s finding; using many hints without spending enough time is not positively correlated with learning probably because students did not exert enough cognitive effort to learn the presented information.

As discussed in the main effect section, the more wrong answers students submitted in the past, the less likely they are to get the current problem solving step correct. However, the positive coefficient for the interaction between the hint requests, \( N_{hint} \), and the number of wrong answers submitted in the past, \( N_{wrong} \), indicates that the main effect from the wrong answer submission changes depending on the hint use of students. As students use more hints, as long as they invested enough time in using them, the negative effect of submitting wrong answers seems to be decreasing. When the other explanatory variables including the number of hint requests are held constant at their mean values, the probability of correct first try decreases as the number of incorrect submissions occurred in the past increases. However, when students used more hints, the probability of correct first try increases because the positive coefficient of the interaction term balances out the negative main effect from the number of wrong answers (see Figure 4A).

Similarly, the negative coefficient for the interaction between hint requests, \( N_{hint} \) and correct answer submissions, \( N_{correct} \), appears to cancel out the positive main effect from the number of correct answer submissions. As shown in
the solid line in Figure 4B, the main effect for the number of correct answers students submitted in the past is associated with a higher probability of getting the current problem solving step correct without using hints available in the learning environment. However, as students used more hints while solving relevant problems in the past, the probability of correct first try does not increase as much (see the dotted line in Figure 4B). These results seem to suggest the conditions under which the use of hints in Andes Physics Tutor is most beneficial to students. Students are more likely to resolve the current problem solving step successfully (1) when they use hints in conjunction with their incorrect answers; (2) when they do not use too many hints on one problem; and (3) when they invest enough time trying to understand the information presented in hints.

![Figure 4](image)

*Figure 4. Interaction effects between $N_{hint}$ and $N_{wrong}$ (A) and $N_{hint}$ and $N_{correct}$ (B)*

**Conclusion**

In this study, the log files capturing how college students used a computer-based tutor while trying to solve difficult physics problems were carefully analyzed to build a statistical model that estimated the likelihood of their successful problem solving from their problem solving history observed in the past. The logistic regression model developed in this study was able to predict the problem solving performance of students with an error rate of about 30%. Various model evaluation methods, such as AUC/e-statistic and 10-fold cross-validation, suggest that the developed statistical model was fairly robust and would be able to maintain a similar predictive accuracy with unseen cases in the future.

Koedinger and Aleven (2007) point out that it is critical to balance giving and withholding information or instructional supports in a computerized learning environment in order to maximize student learning. If learning environments give students too much information prematurely, students may not be able to acquire a schema from learning tasks because they do not exert enough cognitive efforts (Kapur, 2008; Schmidt & Bjork, 1992). On the other hand, if learning environments do not provide instructional supports, academically weaker students are likely to flounder, waste their time with no success, and get frustrated with failure. In most computer-based learning environments, simple heuristics or the learner’s discretion is used to determine when to provide instructional supports. However, simple heuristics, such as giving hints or feedback after students fail to resolve a learning task $N$ times, would be unlikely to find the right moment for providing instructional assistance that can maximize the learning outcome of students. Likewise, providing instructional supports on the learner’s demand may not lead to improved learning because previous studies found that especially novice learners do not possess enough metacognitive ability and prior knowledge required to determine the right moment to ask for help (Clark & Mayer, 2003; Lawless & Brown, 1997). The findings of this study may suggest that it is possible to develop a computerized learning environment that can provide instructional assistance based on the estimated probability of successful problem solving. A logistic regression model built from the log files created by students in the past semesters can be
used to estimate a probability for students in the current semester to successfully solve physics homework problems at their first attempt. Instructional scaffolding can then be provided only when the estimated probability is less than a certain threshold value. Further research is needed to investigate how to determine an optimum probability threshold, and to examine whether this approach can in fact improve students’ problem solving performance.

The statistical model reported in this study was developed from the log files created by all students enrolled in the general physics course taught at USNA. Since Andes Physics Tutor seemed to be used as an add-on to regular class activities, such as lectures and written problem sets, some students used the computer tutor more actively than others. Therefore, analyzing the subset of data including only students who used Andes Physics Tutor more actively might lead to the development of a more accurate and robust statistical model. Similarly, the current statistical model considered all previous problems that were relevant to the current problem solving step in question, which resulted in a rather long data pre-processing time. By limiting the data pre-processing to most recent, instead of all, problems relevant to the current problem solving step, it would be possible to cut down on the data pre-processing time while maintaining similar discriminatory accuracy and robustness in the final prediction model. Further research is required to determine the number of recent relevant problems that can maximize the predictive power of the statistical model while minimizing data pre-processing time.

The fact that there are many different ways to solve physics problems, especially in the college level, might have made it more difficult to build a robust statistical model that can make accurate predictions on the problem solving performance of students. Therefore, it would be interesting to build a predictive statistical model in other domains, such as elementary algebra, that are less complicated than college physics, and compare its accuracy and robustness to those obtained from more complicated knowledge domains.

Also, it would be valuable to investigate whether more advanced statistical and data mining algorithms can be used to build a better predictive model. The logistic regression used in this study belongs to a family of statistical analysis models called Generalized Linear Models (GLM) which assume a linear relationship between explanatory and response variables. Although linear models such as logistic regression have advantages over other approaches in terms of interpretation and inference, it may have weaker predictive power because the linearity assumption is almost always an approximation. There are several ways to address this issue. The simplest approach would be to extend the logistic regression model by adding extra explanatory variables by raising each explanatory variable to a certain power. Employing Generalized Additive Models (GAM) might be another approach that can relax the linearity assumption of GLM. GAM provides a general framework for extending a standard linear model by allowing non-linear functions of each of the explanatory variables while maintaining additivity (James, Witten, Hastie, & Tibshirani, 2013). Similarly, more advanced data mining algorithms that involve non-linear transformation of explanatory variables, such as Support Vector Machines (Cristianini & Shawe-Taylor, 2000), might be used to build a better predictive model.

Another line of research worth pursuing is to investigate the effect of problem difficulty. Obviously, students would take more time, make more mistakes and request more hints when they were working on more difficult problems. However, the difficulty of problems was not considered in the logistic regression model developed in this study. More research is required to find out the best way to incorporate the problem difficulty into the statistical model building process, and to examine whether it can improve the predictive power of the resulting statistical model.

References


The Comparison of Solitary and Collaborative Modes of Game-based Learning on Students’ Science Learning and Motivation

Ching-Huei Chen*, Kuan-Chieh Wang and Yu-Hsuan Lin

Department of Industrial Education and Technology, National Changhua University of Education, No. 1, Jin-De Road, Changhua, Taiwan // chhchen@cc.ncue.edu.tw // yorainchu@yahoo.com.tw // yousyuan@yahoo.com.tw

*Corresponding author

(Submitted January 21, 2014; Revised August 7, 2014; Accepted September 3, 2014)

ABSTRACT

In this study, we investigated and compared solitary and collaborative modes of game-based learning in promoting students’ science learning and motivation. A total of fifty seventh grade students participated in this study. The results showed that students who played in a solitary or collaborative mode demonstrated improvement in learning outcomes. No significant difference was found between the two modes in terms of learning motivation. However, our qualitative results revealed critical insights into our understanding of the effectiveness of game-based learning. Game-based learning encourages students to explore science concepts explicitly and mindfully; furthermore, learning supplemented with collaborative learning can enrich the learning experience and collective problem solving that brings students to the next level of learning. The findings from this study may provide guidelines for what works in the context of game-based learning.

Keywords

Game-based learning, Science learning, Motivation, Collaboration

Introduction

The Program for International Student Assessment (PISA) reported that students’ attitudes toward and literacy levels in science decreased tremendously over recent decades. Science educators and researchers have called attention to the need to promote knowledge of what it means to do science rather than rote memorization to solve science problems. Thus, it is critical for science learning to be a meaningful and engaging experience so that students can easily draw connections between conceptual changes in science learning and theoretical changes in the science community. With advances in modern technology and the need to engage students in knowledge inquiry, educational games as tutor, tool, and tutee have increasingly been used as cognitive tools to support science learning. As a result, the guidelines on the effective use of educational games to promote individual knowledge construction and development of social skills such as communication and collaboration have been suggested by the researchers (e.g., Gee, 2003; Li & Tsai, 2013; Prensky, 2001). However, the desired motivational outcomes and learning gains do not occur naturally while playing games (Gros, 2007; Virvous, Katsionis, & Manos, 2005; Salen, 2007). Given the increasing popularity of using games for instructional purposes, researchers have sought to identify factors within instructional settings that can maximize the effectiveness of this instructional medium (Dickey, 2006; Gedik, Hanci-Karademiric, Kursun, & Cagiltay, 2012; Villalta et al., 2011; Wolfe, 2001). To better understand the instructional settings that can make contribution on the game-based learning, this study aims to investigate the effects of solitary and collaborative modes of game-based learning on students’ science learning and motivation. In the next section, we focus on research within the context of game-based learning and computer-supported collaborative learning. Following this, we conduct an evaluation methodology at the level of secondary education. In the final section, we analyze and discuss the obtained assessment results with regard to the posed research questions.

Literature review

Game-based learning (GBL)

The use of computer games has continually increased in educational settings. As advanced technology, games have transformed pedagogical strategies about learning and teaching (van Eck, 2006). Games are considered useful...
instruments for learning specific strategies and for acquiring knowledge. Games can also be used to learn particular content, and they may leave an impression on the learners (de Freitas, 2006; Rieber & Noah, 2008; Ting, 2010). In recent decades, the integration of games as educational tools has been studied at the primary, secondary, and college levels (e.g., Annetta, Minogue, Holmes, & Cheng, 2009; Gros, 2007; Huizenga et al., 2009; Papastergiou, 2009; Watson, Mong, & Harris, 2011) and has been applied in different disciplines, such as mathematics (Lee & Chen, 2009), biology and genetics (Annetta et al., 2009), history (Watson et al., 2011), and social sciences (Cuenca & Martin, 2010), to effectively achieve various educational goals. More recently, the use of ubiquitous educational games that are wirelessly networked has become an emerging teaching strategy in educational settings. Huizenga and her colleagues (2009) found that students who played mobile games attained higher scores on a knowledge test than those who received a series of regular project-based lessons.

Game-based learning (GBL) offers numerous advantages and benefits that can contribute to a variety of learning outcomes (e.g., Dickey, 2007; Gros, 2007; Huizenga, Admiraal, Akkerman, & Dam, 2009). For example, GBL better mimics real-life situations and teaches learners how to directly apply knowledge and how to address real-life problems they may encounter (Chang, Wu, Weng, & Sung, 2012; Lee & Chen, 2009; Mandinach & Corno, 1985; Yang, 2012). GBL is shown to increase students’ motivation and promote effective learning (Dickey, 2007; Huang, 2011; Miller, Chang, Wang, Beier, & Klischm, 2011; Tuzun, Yilmaz-Soylu, Karakus, Inal, & Kizilkaya, 2009). However, some scholars perceive GBL as an informal learning activity and worry that the use of this approach will not enable students to achieve expected learning goals (Gee, 2003; Gros, 2007; Virvou, Katsionis, & Manos, 2005). As a result, GBL is a man-machine tool, and importing and using this tool to teach or learn should involve careful evaluation of both the game itself and the learners’ perspectives on its usefulness and efficiency (Dickey, 2006; Salen, 2007; Tsai, Yu, & Hsiao, 2012).

Nevertheless, the outcomes of previous studies on the effectiveness of GBL have been mixed (de Freitas, 2006; Tsai et al., 2012). Some researchers have found mixed results (Huang, Johnson, & Han, 2013; Yang, 2012), others have found that games facilitate learning (Miller et al., 2011; Tuzun et al., 2009), and still others have found that games may have a positive effect on learners’ motivation (Liao et al., 2011; Virvou et al., 2005). A literature review of systematic reviews and meta-analysis on the cognitive basis of game-based learning, knowledge acquisition through GBL, and effects of GBL on learning motivation concluded that GBL does not affect knowledge analysis but affects the application of knowledge and recommended that more qualitative studies extend our understanding of the nature of engagement and the learner experience in GBL (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). Wouters and his colleagues (2013) conducted a meta-analysis study on the effects of GBL and indicated that games are more effective in learning and motivation when supplemented with other instructional methods such as collaborative learning. This finding echoes a wide variety of GBL studies that recommend that the research community should continue to focus on what works with which instructional or contextual settings and in which context (Dickey, 2006; Gedik, Hanci-Karademirci, Kursun, & Cagiltay, 2012; Ting, 2010; Villalta et al., 2011; Wolfe, 2001).

Collaborative game-based learning (CGBL)

Collaborative learning can generally be defined as an instruction method in which students at various performance levels work together in small groups or pairs toward a learning goal. Compared to individual learning, in which students work individually at their own level and pace toward a learning goal, collaborative learning requires students to verbalize their thoughts, challenge the ideas of others, and collaborate to achieve group solutions to problems (Koschmann, Kelson, Feltovich, & Barrows, 1996). Moreover, students in collaborative learning modes tend to engage in meta-reflection and higher-level cognition (de Freitas & Oliver, 2006; Feltovich, Spiro, Coulson, & Feltovich, 1996; Hummel, Paas, & Koper, 2006). At the same time, collaborative learning can enhance the development of higher cognitive abilities, cultivate the spirit of collaboration, and formulate positive social relationships with others (Sins, van Joolingen, Savelbergh, & van Hout-Wolters, 2008).

The concept of collaborative learning has been applied in different disciplines such as science (e.g., Chang, Sung, & Lee, 2003; Morris, Hadwin, Gress, Fior, Church, & Winne, 2010; Tao & Gunstone, 1999), social science (e.g., Golanics & Nussbaum, 2008; Heo, Lim, & Kim, 2010; Huang, Wu, & Chen, 2012), medical education (Lu & Lajoie, 2008), and English (Ge, 2011). In science learning, researchers have found that collaborative learning supports learning proficiency in terms of engaging students in the active exploration of core science concepts (Hung, Looi, & Tan, 2005; Jeong & Chi, 2007; Saab, Joolingen, & Hout-Wolters, 2007). Different from traditional collaborative
Purposes of this study

In this study, we developed an educational game that intends to facilitate students’ intuitive understanding of science concepts. The purpose of the game was to examine the effects of different instructional settings, namely solitary and collaborative game-based learning, on students’ learning outcomes and motivation. This study should shed lights on our knowledge about how different instructional settings interact with game-based learning and provide further information upon the integration of instructional settings to develop an informatics system that supports knowledge construction and motivation. In summary, this study conducts an experiment to answer three research questions: (1) How do students in the individual mode differ from students in the collaborative mode with respect to acquiring knowledge?; (2) How do students in the individual mode differ from students in the collaborative mode with respect to motivation for learning?; and (3) How is the game perceived/experienced by students in both the individual and collaborative modes?.

Method

Participants and design

The participants of this study were fifty seventh grade students from two classes at a local middle school; their average age was 13.41. The participants had not learned the instruction that would be taught in this study. Students participated in the study as a learning activity connected with their regular instruction. Participating students were randomly assigned to experimental groups: the individual group or the collaboration group. For the collaboration condition, students were randomly paired into dyads. Altogether, there were 25 students in the individual group and 25 students in the collaboration group (i.e., 11 pairs and a small group of 3 members). During the study, all students were blind to the experimental design and unaware of the existence of other conditions.

Materials

In this study, we designed and developed a game-based learning environment to promote students’ conceptual understanding and motivation in science. Situating learning within the game-based environment can unite two or more unrelated concepts in the minds of learners and can subsequently create a new knowledge schema. This design rationale is not only consistent with constructivist ideas about learning but also helps learners combine new concepts with already understood ideas into a new cognitive structure (Davidson, 1976; Falconer, 2008; Rieber & Noah,
We also strive to design our instructional game that translates what experts know as well as what scientists hold abstractly within their minds into a concrete world that students can physically manipulate (Sun, 2010; Tu et al., 2008; Wolfe, 2001). Lastly, students will form mental models about the instructional game world through inquiry and discovery (Reese, 2003).

The educational goals of our game were (1) to understand the effects of force, (2) to understand the types of force, (3) to describe the force equilibrium condition, and (4) to understand the impact force generated by the object(s). The examples of goals, objects and activities embedded in the game are illustrated in Table 1.

<table>
<thead>
<tr>
<th>Learning goals</th>
<th>Objects in the game</th>
<th>Activities in the game</th>
</tr>
</thead>
<tbody>
<tr>
<td>To understand the effects of force</td>
<td>Different sizes of rabbits and carrots</td>
<td>The size of the rabbit will determine the size of the carrot that can be pulled up.</td>
</tr>
<tr>
<td>To describe the force equilibrium condition</td>
<td>The growing position of carrots</td>
<td>The carrot can be pulled up successfully when the rabbit’s force direction is opposite to the growing position of carrot and its strength is greater than the maximum static friction.</td>
</tr>
</tbody>
</table>

As shown in Figure 1, the main interface of the game included: (1) start, (2) manual, (3) ranking, and (4) quit. Our game runs on Android 2.0 and above systems. When the players clicked on Start, they entered the game. The game included three levels (1, 2, and 3), as shown in Figure 2. These levels represented increasing levels of difficulty and contextual feedback for learning. Players in level 1 worked with only one object and moved around the game environment, whereas players in level 2 (as shown in Figure 3) worked with three objects simultaneously. In addition to the increasing number of objects, the minimum score and time for passing each level also became more
challenging between levels. For example, to pass level 1, players needed to reach 200 points within 60 seconds, whereas in level 3, players needed to reach 500 points within 60 seconds to pass. The manual was instructional messages that taught students to understand types of force, specifically contact forces and action-at-a-distance forces (as shown in Figure 4).

Instruments

A science learning performance test assessed students’ knowledge of force and motion was developed and subsequently evaluated by two science-related course instructors to ensure that the items were adequately and appropriately represented. The items included concept definitions, rules, and application. Three additional open-ended questions were developed to ask students to explain the concepts and give examples. The overall average difficulty level of this test was between .40 and .70, as recommended by Ahmanan and Glock (1981). In addition, this test’s discrimination level was less than .40, as also recommended by Ebel and Frisbie (1991).

The motivational survey was primarily adopted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia, & McKeachie, 1991). Wu and Cherng (1992) translated and evaluated the MSLQ among secondary students in Taiwan and developed a revised version of the MSLQ that has satisfactory reliability and validity. This study used and modified their MSLQ, which included six constructs: (a) intrinsic motivation, (b) extrinsic motivation, (c) task value, (d) control of learning beliefs, (e) self-efficacy for learning, and (f) expectations of success. The numbers of questions in each construct were four, four, six, eight, five, and three, respectively. The rating was based on a 5-point Likert scale. The reliability for all subscales according to Cronbach’s α was 0.86, 0.92, 0.89, 0.84, 0.91, and 0.87 respectively.

The last instrument we collected was the participants’ interviews after the completion of the study. A total of seventeen students were selected for the follow-up interviews. The interview questions were semi-structured and questions were (a) “What have you learned from playing this game? Please elaborate or give examples” and (b) “What has benefited your learning experience throughout this study? Please elaborate or give examples”. These two semi-structured questions attempted to help the researchers clarify and gain insights from each student. The semi-structured questions were also used to help the researchers clarify quantitative statistical relationships and served to triangulate the numeric findings. The interviews were audiotaped and translated and transcribed into English for analysis. The first author analyzed the interview data following the procedures of initial analysis, the individual coding of interviews, and the indexing and clustering of emerging themes, meta-themes and patterns as suggested by Miles and Huberman (1994). To ensure the validity of our qualitative results, a graduate student who was trained and experienced in qualitative analysis was asked to verify that the results were correctly presented and reported. Any discrepancies in the conclusion were discussed until a high consensus was reached.

Procedure

The experimental procedure of this study is illustrated in Figure 5. All participants first listened to a 10 minutes training session to familiarize themselves with the research study procedure and game environment. A pretest score of the force and motion was collected prior to the start of the intervention. After the pretest, researchers divided students into two conditions (individual and collaboration) and were led to different classroom. The purpose to separate students to different classroom was to avoid students in different condition interfered with each other. The researchers then handed the students a tablet and instructed them to read instruction materials about force and motion before starting the game. If the participants were assigned to collaboration groups, they had to play the game and discuss a worksheet together. The time allocated for playing the game was approximately 20 minutes, followed by a posttest and a motivational survey. The total research time was approximately 120 minutes. After the completion of the study, the selected participants were invited to be interviewed for another 15 to 20 minutes.
Results

Learning performance

To measure the differences in each student’s learning performance in different group conditions, a one-way repeated measure analysis was performed. The data analysis for student science learning performance was carried out as a repeated measures analysis with two groups (individual and collaboration) as the between-factor and time of measurements (pretest and posttest) as a within-factor. Dependent variables were scores of the science learning performance. Assumptions for the statistical measures were checked using Levene’s test for equality of variances among the groups. A repeated measures ANOVA with a Greenhouse-Geisser correction determined that science learning performance differed significantly between time points, $F(2, 48) = 5.61, p = .032$. It indicated a significant difference in science learning performance change over time in the two groups. Individual ($M = 43.80, SD = 11.20$) and collaboration ($M = 45.40, SD = 14.54$) groups had similar scores on the pretest at the beginning of the instruction. The individual group ($M = 60.01, SD = 14.21$) scored higher than the collaboration group ($M = 56.40, SD = 19.86$) on the posttest. The individual students increased 16% from the pretest to the posttest. In the collaboration group, students’ scores increased 11% on the tests. The effect size of the gain difference between the two groups was .91.

Motivation

Another dependent variable for student motivation was measured in terms of intrinsic motivation, extrinsic motivation, task value, control of learning beliefs, self-efficacy, and expectations of success. We performed the
Bonferroni-adjusted one-way multivariate analyses of variance to analyze if any differences existed between the two groups in learning motivation. The results showed no difference between the groups for all motivation subscales. The means and standard deviations of these measures for the two groups are presented in Table 2.

| Table 2. Descriptive statistics for the individual and collaboration groups |
|-----------------|-----------------|---------------|---------------|
|                 | *M* | *SD* | *M* | *SD* |
| Intrinsic motivation | 3.10 | 0.98 | 3.42 | 1.10 |
| Extrinsic motivation | 3.23 | 0.63 | 3.25 | 1.00 |
| Task value | 3.10 | 0.84 | 3.43 | 1.12 |
| Control of learning beliefs | 3.32 | 0.54 | 3.52 | 0.64 |
| Self-efficacy* | 3.12 | 0.84 | 3.62 | 0.99 |
| Expectancy for success* | 3.12 | 0.79 | 3.93 | 0.84 |

*p < .05.

**Qualitative results**

Our qualitative data mainly involved interview data to answer the research question three: “How is the game perceived/experienced by students in both individual and collaborative modes?”. Data analysis revealed six themes which collapsed into two major themes: GBL encouraging exploration of implicit science knowledge, and CGBL enriches learning experience and collective problem solving. Pseudonyms were used to protect the identities of the participants, and the experiment condition is indicated in parentheses. The examples from the qualitative data we collected exemplified instances that were influenced and facilitated by playing the game. The major findings of the interviews are summarized below.

**GBL encourages exploration of implicit science knowledge**

The interview data showed that the game had a positive effect on helping students make connections between awareness of what was known and unknown. Carey (collaboration) and Henry (collaboration) said, “This game embedded scientific knowledge that is easy to remember” and “This game allowed me to brainstorm and utilized learned knowledge into practice.” In the traditional classroom, students are usually taught from teachers or textbooks. They seldom ask to explore science concepts on their own. Sofia (individual group) expressed, “I like this game because it had interesting content to learn and play. I think I would like to know more about some of the concepts covered inside.” By playing the game, the students had to explore their knowledge background and think about how the game related to the science content. For example, Winnie (collaboration group) said, “I figured out what contact force is from playing this game, and I also learned friction and gravity by pulling radishes in the game.”

**CGBL enriches learning experience and collective problem solving**

Collaboration is critical in genuine educational experiences. The findings of interview data showed that collaborative game-based learning can increase social interaction and promote reconstruction or co-construction of knowledge schema. For instance, Cindy (collaboration group) and Fong (collaboration group) said, “When I did not understand the problems encountered in the game, my partner would give opinions” and “When we failed to pass the level, we would discuss the reasons and come up with other strategies.” Additionally, we found that collaborative game-based learning can support authentic activities and encourage deeper discussion through collective problem solving and the discussion of rich descriptions of science concepts. Winnie (collaboration group), Henry (collaboration group), and John (collaboration group) expressed how they benefited from their peers: “I enjoyed collaborative learning because you can discuss and interact with others”; “Collaborative learning allows us to help each other and discuss difficulties when encountering problems”; and “I like collaborative learning. For example, my partner had thought about the force direction, which I did not.” These students built relationships with others as they struggled, handled failure, and eventually worked to master the problem presented. Throughout this process, students were not only critically thinking about their strategies but also developing their cognitive abilities. As students tried out various
strategies and assessed the outcomes, they learned to manage problems and understood the importance of working together.

However, some students in the collaborative game-based learning group had serious reservations about collaboration when playing the game with others. Ken (collaboration group) expressed, “I dislike collaboration in the game playing because my partner was too assertive and slow.” Sometimes students just do not get along because their personalities clash. This seems to be even apparent in game-based learning because games are highly interactive, more fun and thrilling, and addictive; thus, certain personality traits will be more adversely affected by playing (Anderson, 2004; Gentile, Lynch, Linder, & Walsh, 2004; Markey & Markey, 2010). Shane (collaboration group) also said, “We made a rule, each one of us, to take turns playing the game. We did not discuss at all.” We often see that the disadvantages of collaborative learning are that some members of the group do not work well together, do not like to work with others, and argue with others. It should be noted that that certain personality traits may lead to an unproductive group. In addition, students’ cultural characteristics may affect their collaborative learning. A past study found that Asian students tend to work cooperatively rather than collaboratively (Zhang, Peng, Hung, 2009).

As a result, collaborative game-based learning may present serious challenges for students in this highly competitive culture.

**Discussion and implications**

The findings of this study indicated that both individual and collaboration groups did not perform differently in the posttest. Aligned with Meluso et al. (2012) and van der Meij et al. (2010)’s findings, this study found the students who played collaboratively did not differ from those who played individually in terms of learning outcomes. Nevertheless, both groups made significant improvement from the pretest to the posttest, which indicates that gaming experience enhanced their overall learning. Our qualitative data offers detailed information on the effects of game-based learning on students’ learning. The results revealed that the game can encourage exploration of implicit science knowledge. While much of human conceptualization involves implicit metaphorical projection from one domain of understanding to another, a game-based learning environment that includes a list of possible sources for science learning can possibly promote conceptual understanding.

In regard to students’ affective perspectives toward learning through playing the game, the results showed that students who played in solitary or collaborative modes had no differences in learning motivation. While no differences were found between the individual and collaboration groups on the learning performance or motivation measures, our qualitative data uncovered veil of collaborative game-based learning that further extends our instructional practices with educational games. In the light of this study, collaborative game-based learning allows students to re-construct or co-construct knowledge schema and encourages collective problem solving and discussion of rich descriptions of science concepts. While collaboration activity in game-based learning can help students solve problems together with diverse perspectives, a positive group dynamic for collaborative learning is necessary to function at its highest efficiency. Conflict between individuals can diminish or stall a group’s ability to work together, which raises a significant problem when personality traits or cultural backgrounds are too diverse to allow for fully formed conflict resolution skills. Mismatched personalities can cause unsatisfactory collaborative learning.

Game-based learning in educational settings has become a common practice. Many of our students are digital natives who already use or play games on their own to understand their world and to learn, often in nontraditional ways, outside of formal education structures. When games are properly used and deployed, we can improve learning, and we can scale up this improvement. This study adds to the existing literature that suggests that games can serve as transformative technological powers for learning and confirms previous studies that found that game-based learning can facilitate learning, particularly conceptual understanding and building (Miller et al., 2011; Tuzun et al., 2009). To extend the current research on game-based learning, we collected qualitative data to extend our understanding of the nature of engagement and learner experience in GBL. It is suggested that the game-based learning approach encourages students to explore science concepts explicitly and mindfully. Although in this study we found that individual and collaboration modes are both beneficial to science learning, our study revealed that game-based learning supplemented with collaborative learning can enrich the learning experience and collective problem solving that brings students to the next level of learning.
Limitations and future research

Modified replication of this study is necessary with the use of alternative designs such as cohort studies or control-group designs so that possible effects of game-based learning can be ensured. Although our game was designed to promote conceptual understanding in the science domain, it is hoped that such instructional approach can transfer learning and support cognitive development in other content domains. In the solitary mode of game-based learning, future study should provide adequately evidence on the interrelationships between motivation, attitude to learning, and learning outcomes. While the recent research has achieved great strides in integrating models of adaptive learning with technology, game environments adapting to individual needs and abilities are still in its infancy. For collaborative game-based learning, future study should investigate individual accountability in addition to group goals in the collaboration process. Moreover, the dialogues of the collaborative pairs could be recorded to provide a more comprehensive understanding of the effective elements of collaborative game-based learning settings. Research investigating game didactics or group assignments is also valuable to advance our knowledge about collaborative game-based learning. The future study should also consider to use larger sample sizes to detect possible effects of various forms of game-based learning.

References


Meluso, A., Zheng, M., Spires, H. A., & Lester, J. (2012). Enhancing 5th graders' science content knowledge and self-efficacy...


The Role of ICT Infrastructure in Its Application to Classrooms: A Large Scale Survey for Middle and Primary Schools in China

Chun Lu¹, Chin-Chung Tsai² and Di Wu*¹
¹National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China // ²Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan // luchun@mail.ccnu.edu.cn // cchtsai@mail.ntust.edu.tw // dr.wudi@gmail.com

*Corresponding author

(Submitted February 27, 2014; Revised June 19, 2014; Accepted August 6, 2014)

ABSTRACT
With the ever-deepening economic reform and international trend of ICT application in education, the Chinese government is strengthening its basic education curriculum reform and actively facilitating the application of ICT in education. Given the achievement gap of ICT infrastructure and its application in middle and primary schools between urban and rural areas, a divide exists between the needs of constructing ICT infrastructure and the patterns of promoting ICT application in education. This paper reports a survey conducted on the infrastructure and application of ICT in middle and primary schools in urban areas (city and county) and rural areas in China based on their demands for promoting ICT in education. The survey focused on 2,168 middle and primary schools, including 717 schools in cities (33.1%), 487 schools in counties (22.5%) and 964 schools located in rural areas (44.5%). Based on stepwise regression analysis, it was revealed that ICT infrastructure had different influences on its application for schools in urban and rural areas. Schools’ proportion of ICT aided courses in counties, and utilization of multi-media classrooms in rural areas may be more associated with the infrastructure. These conclusions would be particularly useful for policy-makers in Asia. Moreover, the analysis model could highlight some areas where improvement plans could be implemented to reduce the digital divide. However, ICT infrastructure was found to play an insignificant role in “utilization of multi-media classrooms” or “proportion of ICT aided courses” in city schools, and thus its role in city schools should be reinterpreted.

Keywords
ICT application in education, ICT infrastructure, Large scale survey, Middle and primary schools, Factor analysis, Stepwise regression analysis

Introduction

The development of ICT in education

The rapid development of ICT has brought both challenges and opportunities to middle and primary schools, particularly as one of the important factors that innovates education by providing equal learning opportunities. Developed countries have always attached great importance to the application of ICT in education. For example, in the United Kingdom, the government spent £2.5 billion on educational ICT in 2008 to 2009 (Yilmaz, 2011). Several surveys have been carried out to investigate the factors related to the use of ICT in teaching and learning by teachers (Back, Jung, & Kim, 2008; Türel, 2011). Integrating computer technologies into education requires successful development of ICT infrastructure(Sadegül, 2006). Depending on the context such as national ICT policy, the implementation process proves to be complex as it is influenced by various agents at different levels and scales. In addition, the schools’ infrastructure and application of ICT have become more visible when examining the development of ICT in education.

ICT development in China

China is currently experiencing rapid economic growth, while the government is trying to eliminate the education divide between rural and urban areas during the process of urbanization and industrialization. It has been found that disparities in access to education between rural and urban areas are the major cause of educational inequality in China (Qian & Smyth, 2008). The Chinese government has consistently given priority to education development, adapting to the international trend of educational reform and ICT application in education (Zhang, Fang, & Ma, 2010). The central government should increase educational transfer payments to undeveloped provinces, and establish a financial system governed by a provincial institution (Yang, Huang, & Liu, 2014).
In China, middle and primary schools are divided into city schools, county schools, and schools in rural areas according to the administrative division. For local or national education administrative departments, in order to equitably and reasonably distribute the funding for ICT infrastructure, they need to compare different demands for ICT infrastructure in different regions (cities, counties and rural areas). Most city schools could offer ICT courses for all students. At least half of the classrooms are equipped with multimedia projectors to support and promote the utilization of digital technologies for learning and teaching, while some of them provide various digital instruction facilities. Besides, various courses can utilize information technology tools to assist teaching in city schools. However, county schools offer ICT courses for the third or higher grade students. The equipment for teachers to use consists mainly of multimedia classrooms, but some teachers of subsidiary courses, such as physics, chemistry, geography and history, cannot use information technology for teaching yet because of the lack of multi-media classrooms and digital instruction facilities. Only some schools in rural areas could offer ICT courses, and only a fraction of schools have multi-media classrooms, and these tend to have a poor operating environment and low utilization rates. Due to the lack of a network environment and hardware facilities, the equipment is not frequently applied in everyday teaching.

**Previous research on ICT infrastructure**

In the past, the relationship between infrastructure and application has been examined in some studies. ICT infrastructure measures the perceived availability and suitability of the ICT tools such as hardware, software and peripheral equipment provided in the school (Vanderlinde & Van Braak, 2010). It also refers to the availability of equipment, software, Internet access and other similar resources in the school (Pelgrum, 2001). A model has been built and tailored to the characteristics of public schools in a developing country (Solar, Sabattin, & Parada, 2013).

ICT equipment and access provided by the school is a key factor in ICT integration. The limitation of the access to technology is not only for computers but also for projectors, resources and other equipment. It has been found that laptop and tablet computers and mobile phones are increasingly considered as useful in education (Prensky, 2001). A literature review of UK research on the topic states that in order for a specific device to be used optimally, certain technical and organizational conditions must be fulfilled (Savill-Smith, 2005). It is necessary to clarify the demands for ICT infrastructure in education based on the current development status to remedy the disparities of regional development and reduce excess investment.

Most previous studies have mainly focused on the benefits brought by a single type of facility or equipment for learning and teaching. However, relatively fewer studies have been conducted to explore the relationship between the infrastructure and application of ICT in developing areas or even depressed areas on a nationwide scale. Therefore, this paper intends to study ICT application in primary and middle schools in cities, counties and rural areas, and to explore the correlation between the application and infrastructure of ICT so as to identify the types of ICT infrastructure to be increased and constructed in different regions in national or provincial plans in education.

**Research purpose and research questions**

Firstly, the main purposes of this study were to develop a model that can be used to identify Chinese middle and primary schools’ ICT application based on the proposed theoretical framework. Secondly, the study aimed to explore the relationship between schools’ ICT application and the development of ICT infrastructure in different regions. Besides, the effects of ICT infrastructure as the predictor of ICT application were investigated. Specifically, the study addressed the following questions:

- Do middle and primary schools in cities, counties and rural areas display grade differences in ICT application, such as “utilization of multi-media classrooms” and “proportion of ICT aided courses”?
- What are the relationships between Chinese middle and primary schools’ ICT infrastructure and their application of ICT?
- Through stepwise regression analysis, can the ICT infrastructure indexes of middle and primary schools be used to make significant predictions about their ICT application?
**Method**

In this large-scale survey, around 4,500 middle and primary schools in 300 districts of 31 provinces in China were surveyed to gather data about their construction of ICT infrastructure and application in education. The research group randomly selected no less than 10 sample districts in each province. The Departments of Education in each province were asked to nominate a provincial coordinator. All middle and primary schools in the sample districts were asked to complete the questionnaire. The leader and head teachers of each school completed the questionnaire according to the current situation of ICT at the whole-school level and sent it back to the provincial coordinator. Finally, the coordinators in each province sent back all the questionnaires to the research group.

In the questionnaire, questions about ICT infrastructure mainly related to the proportion of students attending ICT courses, the number of multi-media classroom seats per student, types of multi-function classroom, educational satellite receivers, and types of digital instruction facilities. The level of ICT application was evaluated according to the utilization of multi-media classrooms and the proportion of ICT aided courses.

A series of ANOVA analyses was employed to examine the infrastructure and application of ICT in the different regions. Subsequently, a stepwise regression model was built according to the analysis of the data collected. Based on the prediction results and the analysis of the model, some suggestions on improving the ICT infrastructure of schools in different regions are provided to improve the ICT application at the national level.

**Sample**

Samples gathered for this research are 2,168 questionnaires of middle and primary schools over 31 provinces in China, and the return rate is 48.2%. These samples were selected randomly, and the scale of the samples meets the standard for correlation and stepwise regression analysis. Among all the valid samples, 717 were schools in cities, accounting for 33.1% of the total, 487 schools were located in counties, constituting 22.5% of the total, and 964 schools were in rural areas, making up 44.5% of the total. The average student number of all the 2,168 schools was 856.09. Of the 2,168 schools, 42.7% (1,064) had over 1,000 students, 29.7% (740) had 500 to 1,000 students, and the remaining 27.6% had 100 to 500 students.

**Instrument**

We developed an integrated model that allowed us to investigate whether ICT infrastructure efforts influence the application of ICT for Chinese middle and primary schools in cities, counties and rural areas (see Error! Reference source not found.).

By gathering national or provincial domestic preliminary research specified in the *China Education Yearbook* about ICT infrastructure in middle and primary schools, a questionnaire was designed based on the comparison of trends and the implementation of this year. The questionnaire for middle and primary schools in *Monographic Research Reports on Construction and Application of Educational Informatization in China* (Group, 2010) was also considered.

Before the questionnaire was implemented, it was sent to the relevant experts including education experts, educational IT academics, school leaders and education officials by e-mail for three rounds of discussion. During the first round of discussion, the 15 experts who participated gave advice on the structure of the questionnaire and the coverage area of the problems. After supplementing the content of the questionnaire according to the experts’ suggestions, the second round of consultation was held. School principals and administrators were consulted about the readability of the questionnaire and the effectiveness of the questions. According to the feedback, the questionnaire was modified again and experts in the field of educational information technology and provincial education administrative departments were consulted before it was then finalized.
Two sections from the large survey were analyzed in this study: (1) ICT infrastructure, and (2) ICT application. We developed an integrated model to investigate whether ICT infrastructure efforts influence the application of ICT for Chinese middle and primary schools in cities, counties and rural areas.

![Diagram]

**Figure 1.** Indexes of ICT infrastructure and application in the model

Although a range of infrastructure indicators could be used to represent a school ICT infrastructure, specific items were chosen according to the recent research and common practice in China. For example, multimedia classroom represents one of the basic ways of conventional e-learning implementation (Zhao & Jiang, 2010). Therefore, “utilization of multi-media classrooms” and “number of multi-media classroom seats per student” were considered in our items. The main places where teachers integrate IT into classroom teaching were the computer classroom, multimedia classroom and multi-function e-classroom (Yeh, Chang, & Chang, 2011). With well-equipped multi-function classrooms, students and teachers can undertake a range of personal and professional learning activities. “Type of multi-function classroom” is therefore considered. Microcomputers, printers, VCRs, camcorders, satellite receivers, and computer modems have become available for use in Chinese education. Schools should make efforts to improve both the multimedia hardware and software devices to establish well-equipped multimedia classrooms and libraries with comprehensive content of resources and courseware (Dai & Fan, 2012). It is also necessary to consider the use of “digital instruction facilities” in basic education. Computer network technology and satellite digital technology have been widely used in teaching in schools and in the management of educational technology of national and local governments (Liu, Cheng, & Liu, 2010). “Educational satellite receivers” are also considered. In middle schools, the level of ICT literacy differed between students who took an ICT subject and those who did not (Hyeoncheol, Soonyong, Jamee, Hongrae, & Junghhee, 2011). Middle schools need to offer ICT as a subject to teach ICT to all students (Aoki, Kim, & Lee, 2013). The primary schools in Taiwan offer ICT courses for third and higher grade students (Lin & Liu, 2010), but only some of the middle and primary schools in China could offer ICT courses for most students. Moreover, according to the *Compendium of Chinese ICT course in Middle and Primary Schools*, the ICT courses in those schools comprise no fewer than 68 class hours per academic year, and the courses performed on computer are no less than 70% of the total lessons (MOE, 2010). Thus, “proportion of students attending ICT courses” in China could reflect the construction condition of multimedia classrooms and computer classrooms at the school level. The measurement of the ICT level of schools indicates how well prepared schools are for using ICT and how much they are using it (Aoki et al., 2013). The use of ICT should not be limited to certain subjects, such as ICT and computer courses, but should also be used in some main and subsidiary courses, such as...
physics, chemistry, geography and history. In our item, “proportion of ICT aided courses” is considered to show the use of ICT in the teaching of all subjects.

At the same time, through a brainstorming session with researchers who had several years of experience in this field and the guide of officials in ICT development in basic education of the Ministry of Education, we identified five independent indexes of ICT infrastructure and analyzed two aspects of a school’s ICT application.

**Indexes of ICT infrastructure and application**

These independent variables were chosen because of their potential impact on ICT application based on previous studies. Although many of them share similarities, they are grouped into different categories, such as ICT courses, multi-media classroom, multi-function classroom and other digital instruction facilities. Without established and verified scales, they work as single item independent variables.

- **Proportion of students attending ICT courses**: ICT courses in middle and primary schools can train students’ abilities to explore problems with the help of ICT, foster their spirit of innovation, and cultivate their practical abilities. Ensuring the quantity and quality of ICT courses has become a part of the development strategy of ICT in education in China.

- **Number of multi-media classroom seats per student**: Together with other digital instruction facilities, multi-media classrooms have become the most commonly adopted modern teaching platform in middle and primary schools. In order to make full use of multi-media classrooms in teaching, a sufficient number of seats must be provided for students.

- **Types of multi-function classroom**: With the widespread use of multi-media technologies in teaching, multi-function classrooms catering to different teaching needs are gradually being used in daily teaching activities in middle and primary schools. The extensive use of multi-function classrooms is definitively to bring innovation and reform to teaching activities and enhance both the depth and width of ICT application in education.

- **Educational satellite receivers**: For schools in counties and particularly in rural areas, educational satellite receivers have become the major way of obtaining the necessary teaching resources due to their lack of access to cable broadband.

- **Types of digital instruction facilities**: Providing diversified digital instruction facilities for teachers and students in the course of teaching will not only satisfy different needs for multi-media applications, but will also enhance students’ interest in making use of such digital instruction facilities in their learning.

This study used a multilevel model to explore the impact of different infrastructure factors on ICT application at the school level. The dependent variables are “utilization of multi-media classrooms” and “proportion of ICT aided courses.” The extensive use of ICT infrastructure in middle and primary schools was mainly reflected in the frequency and scope of the application of ICT infrastructure in daily teaching activities. Hence, in this work, the following two indexes of ICT application were mainly examined:

- **Utilization of multi-media classrooms**: Frequent use of multi-media classrooms indicates that ICT has been regularly used in teaching. The higher the utilization rate of multi-media classrooms, the greater the number of teachers and students who will use the ICT infrastructure to facilitate teaching and learning.

- **Proportion of ICT aided courses**: A high proportion of subjects with ICT aided teaching indicates extensive use of ICT in teaching. Effective use of ICT for education emphasizes the integration of ICT with teaching activities in all or the majority of subjects, and encourages extensive ICT application in teachers’ daily work.

The items for ICT infrastructure and application are shown in *Table 1*.

<table>
<thead>
<tr>
<th>ICT infrastructure</th>
<th>The proportion/ratio of students who attended ICT courses in the recent academic year is ________%.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of students attending ICT courses</td>
<td>The total number of students is__. The total number of seats in multimedia classrooms is__.</td>
</tr>
<tr>
<td>Number of multi-media classroom seats per student</td>
<td>The multi-functional classrooms include ( ): (multiple answers possible)</td>
</tr>
<tr>
<td>Types of multi-function classroom</td>
<td>A. None B. Recording room C. Remote video classroom D. Digital language lab E. Machine room for preparing lessons F. Micro-classroom G. Other</td>
</tr>
</tbody>
</table>
Educational satellite receivers
Have an educational satellite receiver ( ) A. Yes B. No

Types of digital instruction facility
The digital instruction facilities include (multiple answers possible) ( )
A. None B. Computer C. Set Top Box D. Interactive Whiteboard E. Projector F. TV
G. Television with touch screen H. All-in-one machine I. Stenograph J. Multimedia
center console K. Visual Presenter L. Screen for writing M. Other

ICT application
Utilization of multi-media classrooms
The average utilization of multi-media classrooms for teaching is about %.

Proportion of ICT aided courses
The subjects which use ICT for assisted teaching frequently are ( ) (multiple
answers possible)
A. None B. Chinese C. Mathematics D. English E. Physics F. Chemistry G. Politics
H. Geography I. Biology J. History K. Other

Data analysis
In order to draw a diagram of the relationship between the infrastructure and application of ICT in middle and primary schools, the correlations within and between the ICT infrastructure and application factors were analyzed. To examine the differences in these factors across the three regions, an ANOVA analysis was conducted. The ICT infrastructure factors were considered as predictor variables, whereas the ICT application factors were processed as outcome variables. Finally, the important infrastructure factors that influence ICT application in education in schools of different regions were analyzed according to the regression analysis results.

Results
The comparisons of ICT infrastructure and application among middle and primary schools in the three regions
To draw a clear definition among different education development patterns, an ANOVA analysis was employed to compare the ICT application and infrastructure in different regions, as shown in Table 2. Once a significant F-value was obtained in an ANOVA analysis, post hoc tests were widely used to examine the significances of all possible pair-wise comparisons among regions. Table 2 shows the numbers of schools in each region, the mean values of the ICT application and infrastructure indicators in each region, and the comparisons of the post hoc tests. The results of the ANOVA analyses revealed that there are significant differences among regions for the ICT development indicators of “utilization of multi-media classrooms” ($F = 51.15$, $p < 0.001$), “proportion of ICT aided courses” ($F = 43.6$, $p < 0.005$), and for the ICT infrastructure indicators of “number of multi-media classroom seats per student” ($F = 54.92$, $p < 0.001$), “types of multi-function classroom” ($F = 14.89$, $p < 0.001$), “educational satellite receivers” ($F = 89.92$, $p < 0.001$), and “types of digital instruction facility” ($F = 47.06$, $p < 0.001$).

Table 2. ANOVA for ICT application and infrastructure for the three regions

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>F</th>
<th>Post hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of students attending ICT courses</td>
<td>1</td>
<td>717</td>
<td>0.80</td>
<td>24.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>487</td>
<td>0.82</td>
<td>22.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>964</td>
<td>0.79</td>
<td>23.66</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>1.75</td>
<td>(.174)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of multi-media classroom seats per student</td>
<td>1</td>
<td>717</td>
<td>1.36</td>
<td>0.76</td>
<td>1 &gt; 2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>487</td>
<td>1.10</td>
<td>0.71</td>
<td>2 &gt; 3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>964</td>
<td>0.98</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>54.92**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Types of multi-function classroom</td>
<td>1</td>
<td>717</td>
<td>1.24</td>
<td>1.03</td>
<td>1 &gt; 3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>487</td>
<td>1.31</td>
<td>1.08</td>
<td>2 &gt; 3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>964</td>
<td>1.03</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>14.89**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational satellite receivers</td>
<td>1</td>
<td>717</td>
<td>0.18</td>
<td>0.38</td>
<td></td>
</tr>
</tbody>
</table>
Furthermore, a series of post hoc tests (Bonferroni) were performed to compare the differences among regions. The results revealed that the utilization of multi-media classrooms of city schools (mean = 68.84%) is higher than that of county schools (mean = 64.64%) and rural area schools (mean = 57.48%). In addition, the proportion of ICT aided courses in counties (mean = 50.74%) is higher than that in cities (mean = 48.37%) and in rural areas (mean = 46.52%). In terms of ICT infrastructure, city schools are better off than the county and rural area schools.

### Correlation between ICT infrastructure and ICT application

Based on the statistical analysis of all samples, the correlation between the schools’ infrastructure and application of ICT in cities, counties and rural areas is presented in Table 3.

<table>
<thead>
<tr>
<th>Scale</th>
<th>City</th>
<th>County</th>
<th>Rural area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Utilization of multi-media classrooms</td>
<td>Proportion of ICT aided courses</td>
<td>Utilization of multi-media classrooms</td>
</tr>
<tr>
<td></td>
<td>.09*</td>
<td>.08*</td>
<td>.23**</td>
</tr>
<tr>
<td>Number of multi-media classroom seats per student</td>
<td>.31**</td>
<td>.05</td>
<td>.40**</td>
</tr>
<tr>
<td>Proportion of multi-function classroom</td>
<td>.04</td>
<td>.13**</td>
<td>.05</td>
</tr>
<tr>
<td>Educational satellite receivers</td>
<td>.00</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>Types of digital instruction facility</td>
<td>.26**</td>
<td>.11**</td>
<td>.18**</td>
</tr>
</tbody>
</table>

**Note. 1: City schools; 2: County schools; 3: Rural area schools. * p < .001. * p < 0.05.**

For middle and primary schools in cities, it was found that “utilization of multi-media classrooms” was significantly related to selected ICT infrastructure, such as number of multi-media classroom seats per student and types of digital instruction facility ($r = 0.31, p < 0.01, r = 0.26, p < 0.01$, respectively). As for county schools, the infrastructure
index with the highest degree of correlation to the proportion of ICT aided courses is “educational satellite receivers” ($r = 0.48, p < 0.01$). For schools in rural areas, infrastructure related to “utilization of multi-media classrooms” covers “number of multi-media classroom seats per student,” “types of digital instruction facility” and “proportion of students attending ICT courses” ($r = 0.30, p < 0.01$) $r = 0.28, p < 0.01$ $r = 0.10, p < 0.05$, respectively).

In short, there is a certain degree of correlation between infrastructure and application of ICT in schools of all regions. The results of the correlation analysis indicate that the level of schools’ ICT infrastructure is relatively low in counties and rural areas, where the degree of the correlation between infrastructure and application of ICT is higher.

**Stepwise regression estimates for predicting ICT adoption in city schools**

As shown in Table 4, for schools in cities, the indexes that are highly predictive for utilization of multi-media classrooms include the following: “number of multi-media classroom seats per student” ($t = 7.51, p < 0.001$), “types of digital instruction facility” ($t = 5.15, p < 0.001$) and “proportion of students attending ICT courses” ($t = 2.95, p < 0.001$), of which the overall prediction proportion reaches 15%. The results of the statistical analysis revealed that in order to increase the efficiency of multi-media classrooms, schools could increase the number of such classrooms, actively encourage students to study in these classrooms, offer teachers and students diversified digital instruction facilities, arouse students’ enthusiasm for learning with ICT equipment, and increase the proportion of students attending ICT courses. Compared with the proportion of ICT aided courses among all subjects, the utilization of multi-media classrooms depends more on ICT infrastructure, which means that constructing ICT infrastructure is an effective measure to increase the utilization rates of ICT infrastructure such as multi-media classrooms.

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Predicting variables</th>
<th>B</th>
<th>S.E.</th>
<th>B</th>
<th>T</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilization of multi-media classrooms</td>
<td>Number of multi-media classroom seats per student</td>
<td>8.84</td>
<td>1.18</td>
<td>.27</td>
<td>7.51</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>Types of digital instruction facility</td>
<td>2.27</td>
<td>.44</td>
<td>.19</td>
<td>5.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proportion of students attending ICT courses</td>
<td>.11</td>
<td>.04</td>
<td>.10</td>
<td>2.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>46.71</td>
<td>3.77</td>
<td></td>
<td>12.38</td>
<td></td>
</tr>
<tr>
<td>Proportion of ICT aided courses</td>
<td>Types of multi-function classroom</td>
<td>.32</td>
<td>.09</td>
<td>.13</td>
<td>3.45</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>Types of digital instruction facility</td>
<td>.14</td>
<td>.05</td>
<td>.11</td>
<td>2.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>3.98</td>
<td>.22</td>
<td></td>
<td>18.50</td>
<td></td>
</tr>
</tbody>
</table>

Indexes highly predictive of the proportion of ICT aided courses include “types of multi-media classroom” ($t = 3.45, p < 0.001$) and “types of digital instruction facility” ($t = 2.86, p < 0.001$), of which the prediction proportion only reaches 3%. The results show that the ICT infrastructure indexes have slight predictive power for ICT application. Currently, the multi-function classrooms in city schools mainly consist of remote video classrooms and e-lesson preparation rooms. Other types of multi-function classrooms such as microteaching classrooms and IWBs for the training of teaching methods and behavior have also become popular.

**Stepwise regression estimates for predicting ICT adoption in county schools**

As shown in Table 5, for the utilization of multi-media classrooms, average student seat number per classroom ($t = 5.20, p < 0.001$) and types of digital instruction facility ($t = 4.64, p < 0.001$) are highly predictive, with an accumulative prediction percentage of 13%. Regression analysis revealed that the number of multi-media classroom seats per student and the types of digital instruction facility could influence the utilization of such classrooms, which also suggests that the number of multi-media classrooms could not meet teachers and students’ needs for regular teaching use.
Table 5. Stepwise regression between ICT infrastructure and application for county middle and primary schools (n = 487)

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Predicting variables</th>
<th>B</th>
<th>S.E.</th>
<th>B</th>
<th>T</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilization of multi-media classrooms</td>
<td>Number of multi-media classroom seats per student</td>
<td>7.43</td>
<td>1.43</td>
<td>.23</td>
<td>5.20</td>
<td>.13</td>
</tr>
<tr>
<td></td>
<td>Types of digital instruction facility</td>
<td>2.31</td>
<td>.50</td>
<td>.21</td>
<td>4.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>49.54</td>
<td>2.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of ICT aided courses</td>
<td>Educational satellite receivers</td>
<td>3.17</td>
<td>.25</td>
<td>.52</td>
<td>12.95</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>Types of digital instruction facility</td>
<td>.42</td>
<td>.10</td>
<td>.17</td>
<td>4.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of multi-media classroom seats per student</td>
<td>.71</td>
<td>.30</td>
<td>.10</td>
<td>2.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proportion of students attending ICT courses</td>
<td>.02</td>
<td>.01</td>
<td>.10</td>
<td>2.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Types of multi-function classroom</td>
<td>-.43</td>
<td>.19</td>
<td>-.09</td>
<td>-2.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>.29</td>
<td>.85</td>
<td></td>
<td></td>
<td>.34</td>
</tr>
</tbody>
</table>

Indexes which are predictive of the proportion of ICT aided courses of county schools include “educational satellite receivers” (t = 12.95, p < 0.001), “types of digital instruction facility” (t = 4.04, p < 0.001), “number of multi-media classroom seats per student” (t = 2.39, p < 0.001) and “proportion of students attending ICT courses” (t = 2.48, p < 0.001), of which the prediction degree reaches 29%. The results of the regression analysis show that county schools could promote ICT application in all subjects by being equipped with educational satellite receivers and diversified digital instruction facilities, to offer students more seats in multi-media classrooms and more ICT courses. In terms of ICT infrastructure, the predictive effect the index for infrastructure has on the proportion of ICT aided courses is high, indicating that the accessibility of digital resources is important for the development of ICT in county schools.

Compared with other factors, the index for infrastructure is the most predictive for the proportion of ICT aided courses in county schools, suggesting that the construction of education ICT infrastructure could promote ICT application in education effectively in those schools.

Stepwise regression estimates for predicting ICT adoption in rural schools

As shown in Table 6, for schools in rural areas, the predictive indexes that closely correlate with ICT application in different subjects are “number of multi-media classroom seats per student” (t = 9.30, p < 0.001), “proportion of students attending ICT courses” (t = 6.24, p < 0.001) and “types of digital instruction facility” (t = 5.65, p < 0.001). The total predictive degree of all these factors reaches 22%. The factor of infrastructure is more predictive for the utilization of multi-media classrooms in rural schools. Among all the indexes for infrastructure, the number of multi-media classroom seats per student is the most predictive for the utilization of such classrooms, indicating that strengthening the construction of multi-media classrooms can improve the utilization rates in rural schools. However, it also shows that multi-media classrooms in rural schools fail to meet the basic demands of all students. Actually, these schools have reduced chances of using such classrooms for teaching to achieve educational equity.

Table 6. Stepwise regression between ICT infrastructure and application for rural area middle and primary schools (N = 964)

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>Predicting variables</th>
<th>B</th>
<th>S.E.</th>
<th>B</th>
<th>T</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilization of multi-media classrooms</td>
<td>Number of multi-media classroom seats per student</td>
<td>10.08</td>
<td>1.08</td>
<td>.30</td>
<td>9.30</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>Proportion of students attending ICT courses</td>
<td>.19</td>
<td>.03</td>
<td>.18</td>
<td>6.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Types of digital instruction facility</td>
<td>2.23</td>
<td>.40</td>
<td>.18</td>
<td>5.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>27.37</td>
<td>2.56</td>
<td></td>
<td></td>
<td>.70</td>
</tr>
</tbody>
</table>

257
Comparatively speaking, ICT infrastructure in rural areas is more contributive to the proportion of ICT aided courses, so it is important to provide students with enough ICT courses to reach the national standard, and to allocate enough time for students to have hands-on practice. As for the basic hardware facilities for ICT courses, the most important step to take is to increase the number of multi-media classroom seats per student.

With regard to the utilization rate of multi-media classrooms, indexes which are highly predictive include “proportion of students attending ICT courses” \( (t = 9.81, p < 0.001) \), “types of digital instruction facility” \( (t = 4.08, p < 0.001) \) and “types of multi-media classroom” \( (t = 2.87, p < 0.001) \). The accumulative prediction degree of these three indexes reaches 14%. The results of the analysis indicate that a higher proportion of students attending ICT courses together with more diversified digital instruction facilities and multi-function classrooms could advance the progress of ICT aided teaching. The prediction results also show that sufficient and efficient ICT courses could be significantly useful for enhancing students’ IT literacy and popularizing ICT aided courses.

*Figure 2* shows the relationships in the regression analysis results. The ICT infrastructure factors, “proportion of students attending ICT courses,” “average student seat numbers in multi-media classrooms,” and “types of digital instruction facility” could make significant predictions for the ICT application in the three regions. “Types of multi-function classroom” is the predictor for both city and rural area schools. “Educational satellite receivers” is the only predictor to explain the “proportion of ICT aided courses” of the rural area schools.

**Figure 2. Stepwise regression models of ICT application**

In conclusion, improvement of ICT infrastructure plays a limited role in promoting ICT application in ICT aided teaching for schools in cities. However, for the adoption of ICT in teaching in counties and rural areas, such improvement could be greatly advanced.
Discussion

During the past decades, the Chinese government has invested in its ICT education infrastructure increasingly to promote development. The regional disparities of ICT development have become an important indicator of the education balance in China. There is a gap between the ICT application and the investment policy of ICT infrastructure. It is critical to investigate the correlation between infrastructure and application of ICT in middle and primary schools between urban and rural areas.

Our research was to explore the ICT development models and actual needs. “Types of digital instruction facility” is the most important indicator that could predict the “utilization of multi-media classrooms” and “proportion of ICT aided courses” for the three regions. Some Chinese schools have started to explore interactive teaching based upon intellectualization, virtual experiment, and mobile teaching and learning (Ding, Niu, & Han, 2010).

“Proportion of students attending ICT courses” could predict “utilization of multi-media classrooms” for the three regions and “proportion of ICT aided courses” for only rural area schools. It has become very important to empower students to be self-directed knowledge “navigators” on the information super-highway (Skagen, Torras, Blaabjerg, & Hansen, 2006).

“Number of multi-media classroom seats per student” could predict “utilization of multi-media classrooms” for the three regions and “proportion of ICT aided courses” for only county schools. Due to budget constraints, most traditional classrooms are not equipped with LCD projectors. Likewise, access to a projector contributes significantly to teachers’ production of multimedia materials (Hsu & Kuan, 2013). It is important for rural and county schools to offer enough multi-media classrooms in everyday teaching.

“Types of multi-function classroom” could predict “proportion of ICT aided courses” for city and rural area schools. It is possible that if schools provide more types of multi-function classroom, concentrated efforts in fostering peer and technical support would be undertaken. In developed areas, many schools have started to use new instruction software, platform tools and new style IT products, exploring new models and new ways of applying ICT in instruction (Zhang et al., 2010).

“Educational satellite receivers” could only predict the “proportion of ICT aided courses” of county schools. To promote mutual sharing of optimized educational resources between urban and rural areas, rural junior middle schools are equipped with satellite instruction reception sites. Computer network technology and satellite digital technology have been widely used in teaching and management (Liu et al., 2010).

Conclusion

The rapid development of ICT in education in Chinese middle and primary schools has been diminishing the education gap between areas, although it is too early to claim that ICT will eliminate the education divide across the country. To overcome the lack and limitations of the ICT infrastructure, the necessary financial resources should be provided (Goktas, Gedik, & Baydas, 2013). Depending on the type of ownership of physical resources and infrastructure, an important issue is to choose the deployment model appropriate for the educational institution (Despotović-Zrakić, Simić, Labus, Milić, & Jovanić, 2013).

Using a large representative dataset of 2,168 schools in China, a stepwise regression prediction model was constructed to analyze the influence that ICT infrastructure construction may exert on the adoption of ICT. Based on the prediction results of the model and the comparison of schools’ ICT infrastructure and ICT application in cities, counties and rural areas, suggestions for the construction of ICT infrastructure in the three regions are provided for the purpose of further popularizing ICT application in education. Our suggestions for distributing the funding for ICT infrastructure for the different regions are as follows:

For city schools, there are fewer correlation indexes regarding ICT application in all subjects than for schools in the other regions. In general, they should try to make breakthroughs in fields other than ICT infrastructure. In primary
schools, the need for effectively designed materials to use with the hardware and appropriate physical environments have now become prominent in developing countries (Goktas et al., 2013). County schools have reached a certain level in terms of their infrastructure and application of ICT. To achieve balanced application in different subjects, these schools should attach greater importance to the construction of digital teaching resources in every discipline, and try to provide better access to resources for teachers and students.

Rural schools are currently lacking in multi-media teaching environments, especially multi-media classrooms. More educational investment and ICT infrastructure should be allocated to disadvantaged areas. Therefore, the construction of multi-media classrooms could significantly increase the utilization rates of ICT infrastructure for schools in rural areas. Moreover, ICT courses should pay great attention to improving students’ abilities in using ICT technologies and adapting to the ICT learning environment.

There are limitations of the current research, and these issues suggest directions for further investigation. In this research, ICT infrastructure factors are emphasized. However, other researchers may find that additional variables have collective effects which should be examined, such as ICT education resources, ICT policy for education, and ICT integration for teachers. Moreover, ICT infrastructure was found to play an insignificant role in promoting ICT application in city schools. To determine which variables mentioned above are significant predictors for city schools is another challenge for analysis. Future studies can investigate the trend of the digital gap across regions to apply the knowledge from the analysis of ICT development related to the regional digital divide.

Acknowledgements

This work was supported by National Science & Technology Pillar Program during the Twelfth Five-year Plan Period under grant No.2014BAH22F01.

References


Impeding Phenomena Emerging from Students’ Constructivist Online Game-Based Learning Process: Implications for the Importance of Teacher Facilitation

Morris Siu-yung Jong* and Junjie Shang

Department of Curriculum and Instruction & Centre for the Advancement of IT in Education, The Chinese University of Hong Kong, Hong Kong // Graduate School of Education, Peking University, China // mjong@cuhk.edu.hk // jjshang@pku.edu.cn

*Corresponding author

(Submitted March 10, 2014; Revised November 8, 2014; Accepted December 11, 2014)

ABSTRACT

Virtual Interactive Student-Oriented Learning Environment (VISOLE) is a pedagogical approach to integrating constructivist online game-based learning (COGBLe) into formal teaching in school education. This paper reports a qualitative case study on the implementation of VISOLE (in secondary Geography education) in which we probed into the impeding phenomena emerging from the course of students’ learning. Four students—a non-gamer student, a gamer student, an examination-oriented student, and an angry student—were the focal units of analysis. Apart from identifying three impeding phenomena—impromptu gaming, arbitrary gaming, and halt gaming—we observed how the teacher tried to mitigate these phenomena during the implementation process. Besides providing the field with a new understanding of harnessing online gaming in teaching and learning from the student perspective, the study revealed the importance of teacher facilitation in COGBLe.

Keywords

Game-based learning, Educational games, Learning process, Teacher facilitation

Introduction

Supported by today’s online technology, from the perspective of education, the Internet is shifting from an information transmission-and-reception medium between teachers and students, to a constructivist teaching and learning environment. Against this backdrop, the discussion of the educational use of online games (hereafter referred as games) has moved from “sugaring the pill” to providing learners with student-centric learning opportunities (Games & Squire, 2011; Howland et al., 2012). This pedagogical ambition is termed COGBLe—Constructivist Online Game-Based Learning (Jong et al., 2013).

In the recent decade, COGBLe researchers have endeavoured to investigate the ability games to stimulate students’ learning engagement and motivation (e.g., Erhel & Jamet, 2013; Eseryel et al., 2014), provide students with near real-life situated learning experience (e.g., Andrews et al., 2010; Berns et al., 2013), and foster the establishment of learning communities (e.g., Jong, 2014; Kemp & Livingstone, 2006). However, most of the studies have focused largely on transforming games into another type of technology-enhanced constructivist learning environments for informal education, extra-curricular activities, or various short-term research experiments (e.g., Aylett et al., 2008; Gee, 2007; Gee & Shaffer, 2010; Lee et al., 2006; Prendsky, 2010; Shaffer, 2007). There has been less attention on adopting COGBLe in formal education.

We proposed VISOLE—Virtual Interactive Student-Oriented Environment (Jong et al., 2010a), a teacher-facilitated pedagogical approach to integrating COGBLe into formal teaching at school. In our prior preliminary study on harnessing VISOLE in senior secondary education (Jong et al., 2010b), we obtained a significantly positive result in terms of students’ knowledge advancement. On the other hand, via interviews with the students, we learned that a number of unforeseen phenomena which had emerged during the course of VISOLE, impeded their learning process. Nevertheless, the interview data were far from being thorough to give thick description (Geertz, 1973) of how and why these phenomena emerged, how they impeded the students’ learning, and whether the teacher facilitation (either the designated teacher facilitation tasks specified in VISOLE, or any other new teacher interventions) could mitigate these phenomena. Moreover, owing to some practical reasons (see the section of “Our prior work”), the study was conducted in the form of a competition, rather than a formal teaching activity. In other words, VISOLE had yet to enter a “real” classroom.
The present research was a qualitative case study on the implementation of VISOLE in formal curriculum teaching in a senior secondary classroom in Hong Kong. Upon our previous study (Jong et al., 2010b), this research focused on probing into impeding phenomena emerging in the course of VISOLE. We organize the rest of the paper as follows. The two coming sections are a review of recent COGBLe initiatives and a recall of our previous work. After that, we will delineate this study’s research questions, design and findings. At the end of this paper, we will discuss the implications and limitations of our research, as well as giving our concluding remarks and setting out future work.

**Recent COGBLe initiatives**

We categorize recent COGBLe initiatives into two paradigms, namely, *education in games* (EIG) and *games in education* (GIE). The EIG paradigm advocates adopting existing recreational games in the commercial market for educational use. For example, Adams (1998) adopted a popular recreational game, *SimCity*, in his Geography course for equipping college students with urban planning concepts. His proposition was that the simulations in this game could help students understand some real-world conditions of designing a city, and experience successful or disastrous consequences of decision-making in urban planning. The educational use of *SimCity* led to further exploration of the educational potential of some technically similar recreational games such as *SimEarth* for learning of ecology, and *Zoo Tycoon* for learning of zoology (Prensky, 2010). Kemp and Livingstone (2006) harnessed a popular MMORPG (massively multi-player online role-playing game), *Second Life*, to engage university students to participate in various socio-cultural learning activities. More recently, another EIG instance at university, Rankin and Shute (2010) re-purposed *EverQuest II*, another popular MMORPG, to promote learning in the context of second language acquisition.

The above EIG instances, however, could not shed much light on the aspect of adopting COGBLe in formal teaching at school, because the “education” taking place in these instances was rather ad hoc—the learning objectives were set and verified deliberately according to what these games could teach per se. Unlike colleges or universities, school education is driven largely by the statutory curricula with fixed learning objectives (Lai & Lam, 2011; Lam & Yeung, 2010). In reality, it is difficult for school teachers to identify what and how a particular part of a recreational game can be mapped onto a particular learning objective in a particular curriculum (Rice, 2007).

Instead of relying on existing recreational games, GIE researchers developed their “educational games” based upon their pedagogical propositions and articulated with explicit learning contents. For example, *distributed authentic professionalism* is the underlying pedagogical proposition of Shaffer’s (2007) *epistemic games*. Shaffer realized that members of a profession have an epistemic frame—a particular way of thinking and working. Thus, developing students to be members of a particular profession is a matter of equipping them with a right epistemic frame. To accomplish this, Shaffer and his colleagues developed a number of epistemic games that allow secondary students to participate in simulations of various professional communities. *Folklore-based learning*, which is the underlying pedagogical proposition of Lee et al.’s (2006) *problem-based learning game*, suggests that learning takes place in an interactive adventure articulated with a number of problem-solving tasks in a folklore-based story plot. As prototype work, Lee et al. developed a probability-learning game, *Tong Pak Fu & Chou Heung*, to realize their proposition. *Emergent narrative*, which is the underlying pedagogical proposition of Aylett et al.’s (2008) *narrative games*, suggests learning through role-playing in a digital story. The plot of the story in a narrative game emerges from the interactions between players’ avatars and the NPCs (non-player characters) therein. *FearNot*, which is an example of narrative games, was designed to educate primary students against bullying behaviour at school.

Regarding most of today’s education systems, educating formal subject-specific curricula is still one of the most important parts in school education (Lai & Lam, 2011; Lam & Yeung, 2010). Nonetheless, the above GIE instances were not targeted for this purpose. Shaffer’s (2007) epistemic games were aimed at facilitating after-school activities in community centres. Lee et al.’s (2006) problem-solving game was created for conducting a learning experiment on their proposition of folklore-based learning. Aylett et al.’s (2008) narrative games were developed particularly for a campaign addressing children’s bullying problems in European countries. To fully explore the educational potential of games, research on harnessing COGBLe in formal curriculum teaching is a vital area worth further efforts.

Lately, upon the GIE paradigm, the development of *serious games* (e.g., Andrews et al., 2010; Sawyer & Smith, 2008) has become another focus in COGBLe research. Serious games, which refer to the games implemented with the state-of-art gaming technologies, are designed for instruction purpose rather than entertainment purpose (Games
These games are usually composed of so-called “authentic simulations” for providing students with near real-life learning contexts as similar as the ones in the reality (Cannon-Bowers, 2010). However, technological educators (e.g., Clegg, 1991; Peters & Vissers, 2004; Thiagarajan, 1998) have been reminding us of the fact that even high-fidelity simulations can never be exact reflections of the real-world; students usually have problems in making connections between the simulated situations in a game and the referred real-world systems.

Previous scholarship in the field has ignored largely the importance of teachers in the course of COGBLe (either GIE or EIG). In fact, this violates the literature on constructivist education in which teachers should play a vital role. Although constructivist learning theories emphasize an active, self-regulated role for students, it is believed that teachers are still the best at observing when, what and why students encounter problems in the learning process, and scaffolding them to solve the problems (Brush & Saye, 2002; Collins et al., 1989). COGBLe should be no exception.

Our prior work

Drawing on the notions of situated learning (Lave & Wenger, 1991) and intrinsic motivation (Malone & Lepper, 1987), we studied the possibility of integrating COGBLe into formal school education to enable students to learn constructivistly via interactive multi-player gaming. Our proposition encompassed the creation of a near real-life online interactive world modeled upon a set of formal curriculum-related multi-disciplinary domains, in which each student plays a role in this “virtual world” and shapes its development. To realize the idea, we developed Farmtasia—a multi-player interactive online game, and proposed VISOLE—a pedagogical approach to integrating COGBLe into formal teaching at school.

Farmtasia

Collaborating with two Geography professors at our university and two senior Geography teachers in secondary schools, we built the content of Farmtasia upon a multi-disciplinary theme-based module, Agriculture, in the senior secondary Geography curriculum in Hong Kong, involving the knowledge of natural environment, government, economics, technology, production systems, and environmental problems. Apart from the game, we also developed an online knowledge manual as a learning resource for students to look up relevant knowledge in order to solve the difficulties and puzzles arising in Farmtasia. In addition, we built an online blogging platform to help students document their knowledge gained and reflect on their learning experience in the game.

The orchard for rearing horticultural fruits.

The cropland for cultivating vegetables.

The rangeland for keeping livestock.

The Wise Genie, who is a non-player character, will appear in the virtual world for giving some initial hints to players at the beginning of the game. His “showing-up” frequency will decrease gradually, and he will disappear completely starting from Round 4.

Figure 1. Interface of Farmtasia
Farmtasia is a round-based game with 12 rounds in total. Every round (1 hour) equates to six months in the virtual world. Figure 1 shows the interface of the game. Farmtasia features interacting farming systems (cultivation, horticulture, and pasturage) which are modeled by authentic geographical, botanical, biological, and economic models. In the game, each player (student) performs as a farm manager to run a farm composed of a cropland, orchard, and rangeland. He/she competes for financial gain (the quantified gaming outcome) of his/her farm with the other three players who are at the same time running their own farm somewhere nearby in the same virtual world. Throughout the gaming period, they have to formulate various investment and operational strategies to yield both quality and abundant farm products to be sold to the market. They should always keep an eye on the contextual factors (e.g., temperature, rainfall, wind-speed, etc.) of the virtual world so as to perform just-in-time actions, such as cultivating and reaping crops at appropriate times. The full details of Farmtasia can be found in our previous publication (Jong et al., 2010a).

**VISOLE**

We believe that research on harnessing COGBLe in formal school education should not only focus narrowly on creating a game, but also designing a way to help teachers support their students in transforming their gaming experience into learning experience. Based on the theoretical foundations of scaffolding (Vygotsky, 1978), reflection (Dewey, 1958), and debriefing (Crookall, 1992), we proposed VISOLE in which there are three pedagogical phases, as illustrated in Figure 2. Let us use the adoption of Farmtasia in teaching the Agriculture module to elaborate what a teacher and his/her students should conduct in the course of VISOLE. In Phase 1—Scaffolding, through a number of face-to-face lessons, the teacher equips students with “just enough” preliminary high-level abstract knowledge related to natural environment, government, economics, technology, production systems, and environmental problems (as their prior knowledge to the next phase). The activities in Phase 2—COGBLe and Reflection cross over with Phase 3—Debriefing. Phase 2 engages students in playing Farmtasia. All tasks in the game are problematically open-ended. In order to accomplish the tasks, students have to acquire new knowledge on their own from the knowledge manual. In addition, after each round of gaming, they are required to “blog” a journal on the blogging platform to reflect on what they have learned in Farmtasia. In Phase 3, the teacher observes students’ gaming proceedings at the backend, and will extract interesting or problematic scenarios in Farmtasia for conducting case-study-based debriefing with students through a number of face-to-face lessons.

![Figure 2. Three phases of VISOLE](image)

To enable teachers to perform the facilitation tasks in Phase 3, we developed a teacher console connecting to the game server. The game server records all students’ gaming actions in Farmtasia, and teachers can use the console to review all students’ gaming proceedings in the form of video-playback. The full details of VISOLE can be found in our previous publication (Jong et al., 2010a).
Previous evaluation on VISOLE

We conducted a preliminary evaluative study on VISOLE (with Farmtasia) (Jong et al., 2010b), involving 16 Geography teachers and 254 Grade-10 students from 16 schools. At that time, we were hard to find teachers willing to “risk” adopting COGBLe in teaching of a senior secondary formal curriculum in Hong Kong (as all senior secondary students are required to take an important public examination which is equivalent to the A-level examination in the UK). Thus, the study was carried out in the form of a competition (as an “after-school” activity). In the quantitative part of the study (via pre- and post-tests), we obtained a significantly positive result on the students’ advancement in the curricular knowledge. Nevertheless, in the qualitative part of the study (via interviewing the students after the competition), we noticed that a number of unforeseen phenomena which had emerged during the course of VISOLE, impeded the students’ learning. For instance, some students were frustrated by their on-going poor performance in Farmtasia. Eventually, they refused to go on playing the game. Some students had no motive in playing the game because they did not realize that “learning through gaming” was an effective way to help them do well in the public examination. Some students (who were “veteran” gamers) developed a number of advanced gaming strategies to lead other students in Farmtasia, irritating some “hard-playing” students (who played and learned hard in the game). We also identified some plausible factors leading to these impeding phenomena, such as the students’ prior gaming experience, examination-orientedness, and negative emotion(s) induced during the course of gaming.

In our prior study (Jong et al., 2010b), we preliminarily discovered some impeding phenomena emerging during the course of VISOLE, but the understanding was superficial. The qualitative data gathered from the student interviews were not in-depth enough to give thick description (Geertz, 1973) of how and why these phenomena emerged, how they impeded the students’ participation, as well as whether the teacher facilitation could mitigate these phenomena. All these unanswered questions motivated us to carry out the present study.

Research design

This study was a piece of qualitative case study research (a single-case with multiple focal units of analysis), focusing on gaining an in-depth understanding of impeding phenomena (if any) emerging in the course of VISOLE in the setting of formal curriculum teaching at school. Specifically, we sought “answers” to the following research questions:
- Do any phenomena, which emerge during the course of VISOLE, impede students’ learning process?
- How do these phenomena impede students’ learning process?
- Why do these phenomena emerge?
- Can teacher facilitation (either the designated teacher facilitation tasks specified in VISOLE, or any other new teacher interventions) mitigate these phenomena? How and Why?

Selection of the case

In qualitative case study research, a researcher has to purposively select a case from which he/she can learn most (Creswell, 2008; Maxwell, 2005). In this study, the most fundamental step was to have a Geography teacher who was both competent and volitional to implement VISOLE (with Farmtasia) in teaching of the Agriculture module. Hence, we set up two criteria for selecting the teacher participant. The first criterion was he/she should be familiar with the pedagogical idea of VISOLE and the technical operation of Farmtasia. The second criterion was he/she should be willing to adopt VISOLE in his/her teaching practice.

The initial selection scope was on the 16 Geography teachers who had participated in our prior evaluative study (Jong et al., 2010b). After reexamining the teachers’ data collected previously, we shortlisted nine of them who fulfilled best the two selection criteria. These data included their gaming logs in the teacher enablement training, access logs of the teacher console, responses to the open-ended interview question “do you want to adopt VISOLE in your teaching practice?.” Four months before this study, we approached these nine teachers. Six were interested, while three were not. Among the six interested teachers, finally, we selected a female teacher, Tina (pseudonym) because of the good match between her teaching schedule of the Agriculture module and our research timeframe.
Other five teachers were not selected because they were not going to teach this module within the coming 10 to 12 months.

Tina was aged 38, with 12-year teaching experience at a Band-2 secondary school. Secondary schools in Hong Kong are categorized into three academic bands based on their students’ academic achievement; Band-1, Band-2, and Band-3 are respectively the top, middle, and bottom. The context of this study was Tina’s implementation of VISOLE (with Farmtasia) in teaching her class of 40 Grade-11 students on the Agriculture module.

**Implementation setting**

There were two 70-minute Geography lessons every week in the students’ normal timetable. Tina’s implementation of VISOLE lasted for seven weeks, as showed in Table 1. It consisted of three scaffolding lessons (Phase 1), one game-trial lesson, 12 rounds of gaming (Phase 2), and four debriefing lessons (Phase 3). The students started playing Farmtasia at the end of Week 2, playing 1 round every 2 to 3 days until Week 6. They played the game at home mainly. Nevertheless, in order to facilitate us to observe their “physical” gaming behaviour, we required them to play three rounds (Rounds 2, 4, and 10) during the lessons respectively in Weeks 3, 4, and 6. Tina conducted the debriefing lessons respectively after Rounds 2, 4, 7, and 12.

<table>
<thead>
<tr>
<th>Week no.</th>
<th>Wednesday lesson (70 mins)</th>
<th>Friday lesson (70 mins)</th>
<th>Round(s) completed at home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Scaffolding Lesson I</td>
<td>Scaffolding Lesson II</td>
<td>/</td>
</tr>
<tr>
<td>Week 2</td>
<td>Scaffolding Lesson III</td>
<td>Game Trial Lesson</td>
<td>Round 1</td>
</tr>
<tr>
<td>Week 3</td>
<td>Gaming Lesson I (Round 2)</td>
<td>Debriefing Lesson I</td>
<td>Round 3</td>
</tr>
<tr>
<td>Week 4</td>
<td>Gaming Lesson II (Round 4)</td>
<td>Debriefing Lesson II</td>
<td>Rounds 5, 6</td>
</tr>
<tr>
<td>Week 5</td>
<td>Public Holiday</td>
<td>Debriefing Lesson III</td>
<td>Rounds 7, 8, 9</td>
</tr>
<tr>
<td>Week 6</td>
<td>Gaming Lesson III (Round 10)</td>
<td>Public Holiday</td>
<td>Rounds 11, 12</td>
</tr>
<tr>
<td>Week 7</td>
<td>Debriefing Lesson IV</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

**Focal units of analysis**

Even for a single-case study, it is unrealistic to investigate everything therein; a specific focus for steering the selection of focal units of analysis in the case is vital (Creswell, 2008). Hinted by our prior evaluative study (Jong et al., 2010b), we purposely selected four representative students in Tina’s class to be the focal units of analysis in this study:

- a non-gamer student who had very little experience of gaming;
- a gamer student who had very rich experience of gaming;
- an examination-oriented student whose studying strategy was to obtain good results in examinations;
- an angry student who was irritated by other students during Phase 2 of VISOLE.

**Data collection and analysis**

**Pre-implementation stage**

This stage took place four weeks before Tina’s implementation of VISOLE, aiming to identify the first three focal units of analysis. The work included:

- **Student questionnaire survey.** We administered a student questionnaire to gather the 40 students’ prior gaming experience and examination-orientedness.
- **Teacher interview.** We interviewed Tina for gathering more information about the class’s background so as to verify the survey data.
- **Collection of students’ prior academic performance.** We collected the students’ academic performance in the previous semester so as to gain more understanding of their academic background.
• **Determination of the first three focal units of analysis.** Two weeks before the implementation, having connected all gathered data and information, we determined the first three focal units of analysis: (i) *Ava* (pseudonym), a female non-gamer student; (ii) *Bert* (pseudonym), a male gamer student; and (iii) *Cara* (pseudonym), a female examination-oriented student.

• **Class Visit.** One week before the implementation, we visited the class to start developing a friendly rapport with the students. Purposively, we chatted with Ava, Bert, and Cara in an informal way so as to gain more understanding about their background and further triangulate the data and information collected previously.

**Implementation stage**

This stage lasted for seven weeks (see Table 1) in which we harnessed multiple data collection means for looking into the students’ participation in the course of VISOLE. Figure 3 shows a summary of the collected data, including *participants’ self-elaborations* (gathered via informal chats, interviews, and think-alouds), *researchers’ observations* (we observed all lessons these seven weeks), and *online artefacts* (students’ gaming proceedings, access logs, and journals, as well as Tina’s access logs of the teacher console). Based on our ongoing iterative data collection/analysis work conducted in this stage, we identified the forth focal unit of analysis: (iv) *Dan* (pseudonym), a male angry student.

![Figure 3. Data collection in the implementation stage](image)

**Post-implementation stage**

We adopted Maxwell’s (2005) qualitative analysis approach (*coding, categorizing, memoing, and contextualizing*) to analyzing the data collected in the implementation stage, and employed Creswell’s (2008) thematic development technique of *theme layering, and theme interrelating* to inductively form “answers” to the research questions. For the verification purpose, after the completion of our preliminary analysis, we met with Ava, Bert, Cara, Dan, and Tina individually again to carry out *member checking* (Creswell, 2008) in which we solicited their comments on the accuracy of the descriptive parts in our findings. Throughout the course of the analysis, we adopted Denzin’s (1983) triangulation framework (*data source triangulation, methodological triangulation, investigator triangulation, and theory triangulation*) to verify the findings from multiple angles.
Vignettes of the focal units of analysis

We adopt Pierson’s (1999) two-tier reporting approach to qualitative case study research for discussing the research findings. This section (the first tier) elaborates on Ava’s, Bert’s, Cara’s, and Dan’s participation in the course of VISOLE respectively with four descriptive vignettes. These vignettes are narratives of the “critical episodes” to develop the themes for answering the research questions. The next section (the second tier) will delineate our answers to the research questions.

Ava

Based on the evidence collected in the pre-implementation stage, Ava was identified as a non-gamer student. She was not interested in gaming. Her tiny gaming experience was just about playing the mini-games (such as Solitaire) bundled in Microsoft™ OS. Figure 4 indicates the rounds that Ava played in Farmtasia and her quantified gaming results. Ava did not play Round 1 and her early gaming results were not good. Nevertheless, she performed much better in the later rounds. The following will spell out Ava’s vignette.

Refusing to start playing the game

In the Game-trial Lesson, after demonstrating the operation of Farmtasia, Tina asked the class to play the practising round. However, Ava did not log in Farmtasia; instead, she just watched the classmate who sat next to her to play the game. Also, according to the gaming logs, she did not play Round 1. In our chat with her before the Gaming Lesson I, Ava told us—
- *When I was watching Tina’s demonstration, I already found the game interface very complicated ... too much information on the screen ... so difficult ... it is quite different from the ones [games] I played before.*

Starting up gaming with peer support

According to Tina’s think-aloud clip emailed to us before the Gaming Lesson I, through the teacher console, she noticed that some students did not start playing Farmtasia or played it obtusely. She deemed that it was because not every student could familiarize themselves with the game operation. She proposed inviting some gamer students in the class to act as *game tutors* for helping “weak-players” to kick start their gaming. Hence, a number of gamer students whom we had identified in the pre-implementation stage were suggested to Tina.
In the Gaming Lesson I, Tina introduced the game tutors to the class, and offered all students two choices. They could either (Choice 1) go on playing Round 2, or (Choice 2) play a new practising round, and then finish Round 2 at home. The ones who had not been familiar with the operation of Farmtasia could seek advice from the game tutors. In the lesson, Ava chose to play the new practising round, and approached a game tutor for teaching her how to start playing the game. According to the gaming logs, Ava began to play Farmtasia in Round 2 at home. On her blog, she appreciated the game tutor who had helped her.

Frustration in the early rounds of gaming

Although Ava started playing Farmtasia, her gaming results in the early rounds were bad (see Figure 4). Before the Debriefing Lesson I, we conducted a retrospective interview with Ava in which we allowed her to review her gaming proceedings in Rounds 2 and 3 with the teacher console. After that, we asked her to give justifications on her gaming acts made in these two rounds. We found most of Ava’s justifications stemmed from her intuitive thoughts about Agriculture, for example—

- **(Why did you continuously command your workers to irrigate the cropland in Round 3?)**
  Ava: Everyone knows irrigation is a basic task that farm workers should carry out every day.

- **(When you commanded your workers to irrigate the cropland, were you aware of the humidity of the virtual world?)**
  Ava: What? ... Where can I check the humidity?

- **(Why did your choose to plant apples but not oranges in your orchard?)**
  Ava: Apples cost more than oranges at supermarket. I will earn more money by planting apples.

As indicated from the access logs, Ava did not read the knowledge manual before the Debriefing Lesson I. She just improvised to play Farmtasia based on her “common sense” and/or daily-life experience. In the scaffolding lessons, Tina did emphasize that Farmtasia was “not so easy,” and reminded the class of paying attention to the gaming context, formulating optimal strategies to operate their own farm, and looking up the knowledge manual whenever necessary. However, Ava did not pay attention to the reminders. As a result, after Round 3, she used up almost two thirds of her initial capital because of her ignorance of the cost of manpower in the game. On top of that, owing to her over-irrigation and unawareness of the climate in the virtual world, all crops died in her cropland. Ava expressed her frustration on her blog—

- *I was so frustrated ... it is unfair; all stuff died in my farm ... I really don’t want to play this game anymore.*

Learning through teacher debriefing

Tina conducted two case studies in the Debriefing Lesson I. We will discuss the second one in Bert’s vignette. The first one was a scenario extracted from a student’s (namely Student X’s) gaming proceedings in Round 2. In the scenario, Student X commanded 20 workers to irrigate his cropland 16 times continuously. Tina first let the class guess the consequences of Student X’s act. After that, she replayed the ending of the scenario (see Figure 5)—all crops died, and the farm workers were out on strike. In fact, prior to the Debriefing Lesson I, Student X and many other students had thought that the strike event was just a “fun” element and would take place sooner or later in all farms. They were not aware of the correlation between the overloading of the workers and the strike. After the Debriefing Lesson I, we chatted with Ava again. We noticed further that her improvised acts in the early rounds were largely because of her lack of prior gaming experience and general perceptions of gaming—

- *I neither have much experience nor interest in gaming ... I thought that gaming should be a sort of entertainment, and an effortless and relaxing activity. I didn’t think it was necessary to invest extra time and effort on equipping myself because of gaming. Now I understand I am not merely playing a game, but participating in a highly authentic virtual world.*
Becoming a proactive game-based learner

Ava started looking up the knowledge manual after the Debriefing Lesson I. She accessed it one or two times before playing each round. We also found that Ava copied a lot of Agriculture-related information from the manual onto her blog. In our later chat with Ava, she elaborated that all these were her “gaming notes.”

In the Debriefing Lesson II, Tina asked the students to share their own gaming experience so far with the classmates who were sitting around. During the lesson, we observed that Ava was quite proactive to discuss the happenings in her farm with the peers. She also approached again the game tutor (who had helped her previously) for additional gaming advice. In our chat with Ava after the Debriefing Lesson II, she told us happily that she had learned a number of good ways to run her farm from the classmates, for example—

- Pesticides should not be used two weeks before harvesting.
- Fertilizing the cropland with the excrement of the sheep and castle is a good sustainable development strategy.

In the later rounds of gaming, Ava became more knowledgeable in playing Farmtasia. Her farm started having income at the end of Round 5 (see Figure 4). Ava went on doing well in Rounds 6 and 7, and thus her accumulated capital was increasing. In the middle of Round 8, a catastrophe, locust attack, happened in the virtual world, causing the death of all the growing wheat in Ava’s cropland. After undergoing this catastrophe, Ava accessed the knowledge manual again, and looked up its causes, precautions, and remedies. Although Ava did not win in the game eventually, she found the course of VISOLE memorable and useful. She wrote on her blog—

- These weeks, I learned so much about Agriculture … Thanks Farmtasia, thanks Tina and my classmates. I learned a lot from you. I really want to play the game again. Can I?

Bert

Based on the evidence collected in the pre-implementation stage, Bert was identified as a gamer student. He had rich prior gaming experience, especially in playing recreational games in the commercial market. During the class visit, we also learned Bert’s nickname, “Game God,” from his classmates. Bert was a high achiever in the “game world”; however, he was a low achiever in terms of academic performance. In the previous semester, his academic rank in the class was the second last. Figure 6 indicates the rounds that Bert played in Farmtasia and his quantified gaming results. Bert only played the first six rounds in which the accumulated capital of his farm was increasing from Rounds 1 to 6. The following will spell out Bert’s vignette.
Being an engaging game-based learner

In the early course of VISOLE, Bert was one of the most engaging game-based learners in the class. As observed, Bert very focused on listening what Tina covered in the scaffolding lessons. Bert’s behaviour did surprise Tina. In the chat with Tina after the Game-trial Lesson, she told us—

- Bert usually sleeps in my lessons. I have never seen him to be so serious in a learning activity.

In the Game-trial Lesson, as expected, Bert was able to get himself familiarized with the operation of Farmtasia without any problems. According to the access logs, he looked up the knowledge manual 12 times between Rounds 1 to 3. He copied a considerable amount of the manual’s content onto his blog for developing his gaming strategy notes. When reviewing Bert’s gaming proceedings, we found that he led other three players from Round 1 till Round 3. In these rounds, Bert usually did the right things at the right time. For example, the irrigation frequency of his cropland was regulated according to the humidity of the virtual world. He also reaped the fruits timely so that he could sell them to the market at a good price. Nevertheless, Bert’s gaming acts had yet to be perfect. He did not pay attention to the market information in Farmtasia. According to his gaming proceedings in these rounds, Bert never looked up the market-value window (displaying the current market values of various farm products). The sowing period and growth cycle of both lettuces and Chinese cabbages are similar, but the market value of lettuces was 1.5 times higher than Chinese cabbages in Round 1. If Bert had cultivated lettuces rather than Chinese cabbages, he could have earned more money.

Enhanced gaming strategies after teacher debriefing

In the Debriefing Lesson I, Bert quite engaged in discussing the two case studies conducted by Tina, especially the second one, because it was about the scenario of his unawareness of the current market values of various farm products in Farmtasia. In the chat with Bert after the lesson, he told us that he did enjoy this kind of teaching and learning. He elaborated—

- I like this activity ... I never mind Tina showing my gaming proceedings in front of the class ... it is an honour in fact, haha.
- The debriefing did inspire me to some new gaming strategies ... I will try these strategies in the coming rounds.

In Round 4, we watched face to face Bert’s gaming as the round was arranged being played in the Gaming Lesson II. As observed, Bert was more aware of the market information in the virtual world. Before making any new financial-related decisions, he would look up the market-value window.
Bert was one of the game tutors in the Gaming Lesson I. As observed, he was quite willing to help the “weak players” start playing Farmtasia. In the Debriefing Lesson II, Bert did not hesitate to share his good gaming strategies with the classmates. In the chat with Bert after the Debriefing Lesson II, he told us—

- *I always share my knowledge of playing recreational games with other gamers via game forums. It is very common in the game world.*

**Deriving an degenerate strategy**

Bert’s accumulated capital in Farmtasia bloomed dramatically from $88,541 (after Round 4) to $209,876 (after Round 5), and then to $400,110 (after Round 6), see Figure 6. After reviewing his gaming proceedings, we found that Bert had discovered an *exploit* of the game and developed a corresponding *degenerate strategy* on the exploit. Exploits refer to loopholes in a game (where players can advance effortlessly), while degenerate strategies are ways of playing a game to ensure victory every time (Salen, 2008).

Bert named his degenerate strategy *cattle-scalping*. He also shared this strategy with the classmates via his blog after the completion of Round 6. Scalping refers to “buy tickets for an event and sell them again at a much higher price” (Longman Dictionary of Contemporary English, 2014). In the last two months (the virtual-world time) of Round 5 and the entire Round 6, Bert did nothing except scalping cattle. In fact, this exploit revealed a “bug” inside the supply-and-demand model in Farmtasia. In real life, the cattle price should drop when cattle are largely available in the market.

In Tina’s think-aloud clip emailed to us before the Debriefing Lesson III, she mentioned that she had observed more and more students using the cattle-scalping strategy to “cheat” in Rounds 6 and 7. In the Debriefing Lesson III, Tina discussed the cattle-scalping scenario with the class, and the fault inside the supply-and-demand model in Farmtasia. As observed, Bert did listen to Tina with close attention. At the end of the lesson, Tina launched two interventions (we will discuss the second one in Cara’s vignette). The first one was to warn the class of not adopting the cattle-scalping strategy anymore in the coming rounds. In the chat with Bert after the lesson, he told us how he had formulated his degenerate strategy—

- *I monitored closely the market information … I observed that the purchasing price of cattle was often lower than the selling price. I thought I should be able to earn money simply via purchasing and selling cattle … it worked, haha.*

**Ceasing to play the game**

Bert stopped playing Farmtasia since Round 7. He even skipped Round 10 which was scheduled being played in the Gaming Lesson I. During the lesson, we observed that he was playing a newly released recreational game. In our interview with Bert after the course of VISOLE, he elaborated why he had ceased to play Farmtasia—

- *Farmtasia is the most interesting “educational” games that I have played so far … I was quite moved in playing this game at the beginning. However, I found it a bit boring after playing several rounds, especially after I had formulated the cattle-scalping strategy … compared to commercial games, Farmtasia is still less exciting.*

Although Bert had stopped playing Farmtasia since Round 7, as observed, Bert did pay close attention in the Debriefing Lessons III and IV. Bert elaborated—

- *Yes, I pretty like Tina’s teaching method in the debriefing lessons. The gaming scenarios extracted in the game were interesting … it was much more interesting than looking at the textbook, blackboard, PowerPoint or even educational video.*

**Cara**

Based on the evidence collected in the pre-implementation stage, Cara was identified as an examination-oriented student. Her most preferred teachers’ pedagogical style was the one that could help her get excellent examination results. In terms of academic performance, she was one of the top-three achievers in the class in the previous semester. In the class visit, Cara told us her parents did have high expectation of her studies—
My parents want me to, like my sister, enter the same university ... they want me to study in the famous Law faculty.

Figure 7 indicates the rounds that Cara played in Farmtasia and her quantified gaming results. Between Rounds 1 and 7, she played the game in an on-and-off manner. Starting from Round 8, Cara resumed playing the game till Round 12. The following will spell out Cara’s vignette.

Reckless gaming

In the Scaffolding Lessons I, II, and III, Cara did pay attention to what Tina covered for the class. Nevertheless, she did not like Tina’s teaching style in these lessons, and doubted whether she would really gain the knowledge concerned through gaming. In the chat with Cara after the Scaffolding Lesson III, she elaborated—

- The knowledge covered in these lessons was too little ... it won’t be enough for preparing me to sit for either the internal or public examination ... I don’t believe I can learn well via playing this game.

In the Game Trial Lesson, Cara was able to get herself familiarized with the operation of Farmtasia. Further, we could tell from the smiles on Cara’s face that she found the game more or less interesting. However, among the first seven rounds, Cara skipped five. She only played the two rounds (Rounds 2 and 4) which were arranged being played respectively in the Gaming Lessons I and II. On top of that, according to her gaming proceedings, she played the game recklessly. For example, in Round 2, she bought a milking machine without keeping cattle in her rangeland. She commanded the workers to irrigate her cropland frequently even though the aqua-index of the soil was very high. She neither worried about the consequences of her acts to the farm nor cared about the gaming results. Before the Debriefing Lesson I, we conducted a retrospective interview with Cara in which we allowed her to review her gaming proceedings in Round 2 with the teacher console—

- (Why did you command your workers to irrigate the cropland 32 times in the second week of July?)
  Cara: Irrigation is the basic work in a farm.
- (Why 32 times?)
  Cara: ... (irresponsive)
- (Why did you only play Round 2, but skip Rounds 1 and 3?)
  Cara: Round 2 was arranged being played in the lesson ... no choice, I had to entertain Tina’s request.
- (Are you frustrated with your bad gaming results in Farmtasia?)
  Cara: Certainly not ... it is just gaming ... it is a game, not an examination.

During the period between Rounds 1 and 7, Cara neither wrote anything on her blog nor accessed the knowledge manual. Figure 8 is a screen shot of Cara’s farm captured at the end of Round 4. She employed 36 workers and
bought 10 heads of cattle. However, she neither commanded the workers to conduct any farming-related work, nor fed the cattle.

Figure 8. A screen shot of Cara’s farm at Round 4

Resuming playing the game

In Tina’s think-aloud clip emailed to us before the Debriefing Lesson III, besides the cattle-scalping issue, she also mentioned that three students stopped playing the game for three consecutive rounds (Rounds 5, 6, and 7) and did not write anything on their blogs. The second intervention launched by Tina during the Debriefing Lesson III was, starting from Round 8, she would award two marks to the students who completed a round and blogged a journal. These marks would contribute to their continuous assessment performance in the semester.

Since Round 8, Cara resumed “gaming” and started “blogging.” Nevertheless, according to Cara’s gaming proceedings from Rounds 8 to 12, after logging in the game, she just kept her farm idle for most of the time. She never read the knowledge manual throughout the course of VISOLE. On her blog, she just wrote down some superficial things, such as—
- I successfully completed Round 8.
- I finished Round 9 already.

In our interview with Cara after the course of VISOLE, we probed into the reason why she had resumed playing Farmtasia, but playing it so recklessly—
- (What made you resume playing the game since Round 8?)
  Cara: For the marks, of course.
- (Would you have resumed playing it if Tina had not launched the “two-mark” intervention?)
  Cara: Guess not.

In the interview, Cara stated that she was not willing to spend time and effort on the gaming activity. During these few weeks, she worked hard as usual to prepare herself to sit for the semester-end Geography examination. Cara conducted self-study on the Agriculture module with the textbook at home. Nonetheless, although Cara did not embrace the “in-the-game” part of VISOLE, she deemed that she had benefited from the “off-the-game” part. She elaborated—
- The content [high-level knowledge] covered in the Scaffolding Lessons I, II, III was useful for my self-study ... I gained some useful prior knowledge before pursuing my self-study.
- Yes, I admit that Tina’s teaching style in the debriefing lessons was more interesting and motivating than her conventional ones.
- I had not quite understood the function of crop rotation when reading the textbook, but I gained a better understanding of this farming technique and its importance after watching the gaming scenarios extracted from Farmtasia in Tina’s debriefing.
Dan

Unlike other focal units of analysis who had been identified in the pre-implementation stage, we selected Dan because of his behaviour in the Debriefing Lesson III. In the lesson, he shouted out angrily that Farmtasia was so unfair. Figure 9 indicates the rounds that Dan played in Farmtasia and his quantified gaming results. Before Dan stopped playing the game (i.e., before Round 8), the operation of his farm had been “on the right track” and his accumulated capital had been increasing round after round. According to the access logs, he read the knowledge manual one or two times before playing each round. Dan stopped playing Farmtasia in Round 8, but resumed playing it in Round 11. The following will spell out Dan’s vignette.

Emergence of anger

Figure 10 shows respectively the gaming results of the four students (including Dan) in Dan’s group among the first seven rounds. Dan led other three classmates from Rounds 1 to 5; however, the accumulated capital of one of the classmates (namely, Student R) bloomed suddenly in Round 6. The same happening took place in another classmate’s (namely, Student P’s) farm in Round 7. At the end of Round 7, Dan’s rank in the group dropped to the third. After reviewing the gaming proceedings of Students P and R in these two rounds, we found they used largely the cattle-scalping strategy to play the game. In the later chat with Students P and R, they revealed that they had learned the strategy from Bert’s blog.

Figure 9. Dan’s participation and gaming results

Figure 10. Four players’ gaming results (Round 1 to Round 7) in Dan’s group
In the Debriefing Lesson III, Tina discussed the cattle-scalping issue with the class. After knowing it, Dan was seething. He bellowed—

- Farmtasia is so unfair, and I won’t play it anymore!

After the lesson, before leaving the classroom, Dan looked at us angrily and said—

- How could you develop such a cheap and nasty game!

Teacher emotional support

Dan did not play Farmtasia in Rounds 8 and 9. On his blog, he expressed his upset feeling about the game. Round 10 was scheduled being played in the Gaming Lesson III. In the lesson, Dan did not turn on his computer, but took out the Geography textbook to read. At about one third of the lesson, Tina approached Dan and chatted with him for a while. After the lesson, we interviewed Tina to learn about what she had talked to Dan—

- (What did you talk to Dan?)
  Tina: I gave him some emotional support … I tried to encourage him to resume playing Farmtasia … I told him that I understood and empathized his feeling about the unfairness, but refusing to play the game would make him miss some valuable learning opportunities … his quitting was not a way to punish the misbehaving classmates … I also related the cattle-scalping issue to the “Lehman Brothers” incident happened in the stock market some years ago … I wanted to remind Dan of the fact that there are always misbehaving people in our society; nevertheless, we should always be a responsible person and try our best on duties. Now, his duty is to go on playing Farmtasia.

- (Why is it important for Dan to go on gaming?)
  Tina: Because VISOLE is a “learning through gaming” activity … just like the case that the locust attack in Round 8. Dan didn’t play this round. He missed the chance to experience this catastrophe in the virtual world. So, in the coming debriefing lesson, he might have less reflection on the deconstruction, as well as precautions and remedial strategies related to this catastrophe.

Before the commence of Round 11, Tina further left a supporting message on Dan’s blog—

- Cheer up, Dan! I trust you will never give up!

Resuming playing the game

Dan resumed playing the game in Round 11. According to Dan’s gaming proceedings in this round, the beginning scene showed the death of all crops planted and livestock kept previously in Round 7. Dan then started commanding some workers to carry out the ploughing, sowing, and irrigating tasks in the cropland. Moreover, he bought some cattle and sheep again, and kept them in the rangeland. Before the end of the game (Round 12), Dan was able to harvest the crops and sell them together with the livestock to the market. This trade brought him good financial gain. In the interview with Dan after the course of VISOLE, we probed into the correlation between his gaming resumption and Tina’s emotional support—

- (Did you read the message that Tina had written on your blog?)
  Dan: Yes … her message was encouraging and supportive … Also, in the chat with Tina, she encouraged me not to give up … as a responsible person, I should finish this game.

- (You resumed playing the game solely because of Tina’s encouragement?)
  Dan: Tina’s encouragement counted a lot … I think another critical reason was Tina warned the class of not using the cattle-scalping strategy again. This is very important. Nobody wants to be cheated.

Impeding phenomena

From the vignettes of Ava, Bert, Cara, and Dan, we identified three themes of impeding phenomena, (i) impromptu gaming, (ii) arbitrary gaming, and (iii) halt gaming, for answering the four research questions of this study.
Impromptu gaming

We define impromptu gaming as—when a student is playing the game, he/she improvises his/her gaming acts solely based on common sense or his/her personal experience. In this study, Ava took this approach to playing the early rounds of Farmtasia. She neither looked up the knowledge manual nor formulated any strategies before playing those rounds. Ava’s gaming acts were actually her instinctive and spontaneous responses to what she encountered in the game. However, those acts were neither meaningless nor in a trial-and-error fashion, but entailed “rational guesses” stemming from common sense or her personal experience in real life, i.e., her intuitive knowledge (Swaak & de Jong, 2001).

VISOLE requires students to analyze and evaluate the gaming context of Farmtasia proactively, and then create optimal strategies to play the game. Farmtasia will automatically “punish” students who take the impromptus gaming approach, by giving them bad gaming results. For example, in Ava’s farm, at the end of Round 3, her initial capital dropped nearly two-thirds owing to her unawareness of the cost of manpower. Failure is a critical precondition for learning to take place in gaming (Prensky, 2010). The “punishment” to Ava, or in Shaffer’s (2007) terms, authentic feedback, was aimed at signaling her that she was not doing well so far; she needed to be more knowledgeable to formulate better gaming strategies in order to better operate her farm. However, Ava interpreted that her failure in those rounds was a sort of unfairness to her. She was frustrated.

Ava was a “newbie” to gaming. Before playing Farmtasia, she had realized that gaming should be entertaining, effortless, and relaxing, rather than taxing and demanding. Ava had never thought that she should have spent extra time and effort on equipping herself merely because of gaming. At the beginning, she ignored Tina’s reminders that everybody should pay attention to the gaming context of Farmtasia, formulate optimal gaming strategies, and look up the knowledge manual whenever necessary. With her intuitive knowledge, Ava improvised to play the game in a free and effortless fashion.

Tina’s case-study-based debriefing (a designated teacher facilitation task in Phase 3 of VISOLE) worked to curb Ava’s impromptu gaming approach by letting her understand that she was not merely playing a game, but participating in a near real-life virtual world. After the first debriefing lesson, Ava began to realize that the undesirable happenings in her farm and the bad gaming results were actually caused by her own impromptu gaming acts. She was provided with a good “justification” why she had to spend time and effort on acquiring new knowledge from the knowledge manual. Moreover, Tina’s new intervention, peer-sharing among classmates, in the second debriefing lesson also made Ava more engaging in playing Farmtasia, in which she also learned more good gaming strategies from the peers. She was transformed from a frustrated game-based learner to a proactive one.

Arbitrary gaming

We define arbitrary gaming as—when a student is playing the game, his/her gaming acts are haphazard or even illogical, without caring about his/her gaming results. In this study, Cara took the arbitrary gaming approach to playing Farmtasia. For example, in Round 4, she employed 36 workers, but did not command them to conduct any farming-related work. Among the first seven rounds of gaming, she only played two; she neither accessed the knowledge manual nor wrote anything on her blog. Cara neither worried about the consequences of her arbitrary acts to the farm nor cared about the gaming results.

Cara’s arbitrary gaming approach made her have no learning progression in Phase 2 of VISOLE. A premise of COGBLe is that, undesirable happenings in games or bad gaming results will lead to discrepancies between what students understood in the past and what they have experienced in the games (Gee, 2007). Further, learning is experience plus reflection (Dewey, 1958). By reflecting on the discrepancies, students will establish new learning goals for what they need to acquire in order to make sense out of those discrepancies. This constructivist process, however, was not witnessed in Cara’s gaming proceedings or on her blog.

Cara was an examination-oriented student, or in Biggs and Moore’s (1993) terms, achieving student. She realized “learning through gaming” was unable to help her reach her study goal, i.e., obtaining a good result in the semester-end examination. Cara was unwilling to spend time and effort on playing Farmtasia; instead, she conducted self-study on the Agriculture module with the textbook.
In terms of resuming Cara’s access to Farmtasia, Tina’s new intervention, participation-based marking scheme (i.e., the “two-mark” intervention), was successful; however, in terms of curbing her arbitrary gaming acts, it failed. The intervention only raised her extrinsic motivation, not achieving motivation (Biggs & Moore, 1993). Cara just kept taking the arbitrary gaming approach to playing the remaining rounds, and wrote some superficial content on her blog. She resumed her gaming and started blogging merely because of the marks which would contribute to her continuous assessment performance in the semester.

Halt gaming

We define halt gaming as—a student refuses to start or go on playing the game. In this study, halt gaming emerged in Ava’s, Bert’s, and Dan’s participation in VISOLE. Ava was unwilling to start playing Farmtasia, while Bert and Dan stopped playing Farmtasia respectively after six and seven rounds of gaming.

Notwithstanding the inclusion of the “off-the-game” elements (scaffolding, debriefing, and reflection), a considerable part of the learning in VISOLE relies on gaming. If students refuse to play Farmtasia, they will have no significant progression in their learning process. Let us use the catastrophe taking place in Round 8 as an instance. The injection of the locust attack was aimed at provoking the students into acquiring the knowledge related to that catastrophe. Regrettably, since both Bert and Dan did not play that round, they missed that constructivist learning opportunity.

Ava, Bert, and Dan halted their gaming due to different reasons. In Dan’s case, after knowing the classmates “cheated” in Farmtasia, he was irritated and upset. Unfairness in gaming always arouses players’ anger (Ferdig & Mishra, 2004). Owing to the anger, Dan stopped playing the game. Concerning Ava and Dan, their halt gaming acts could be explained by Csikszentmihalyi’s (1975) theory of “boredom and anxiety.” When a person is working on a task, if the challenges therein are far below his/her ability, boredom will be provoked; on the other hand, if the challenges are far beyond his/her ability, anxiety will be provoked. Either boredom or anxiety will lead him/her to escape from the task. Ava was a non-gamer. Owing to her lack of gaming experience, she found that Farmtasia was too difficult to play. She was anxious and thus refuse to start playing it. Bert was a veteran gamer. He found Farmtasia “a bit boring” and stopped playing it since he had derived the cattle-scalping strategy. About the same time, he started engaging in another newly released recreational game that he found more exciting.

Tina’s new interventions mitigated Ava’s and Dan’s halt gaming acts, but failed to make Bert resume playing Farmtasia. In Ava’s case, the intervention, peer support by gamer students, worked to assist Ava in starting up her gaming. With the “scaffolding” (Vygotsky, 1978) offered by a game tutor, Ava began to play the game. In Dan’s case, Tina’s new interventions, cease of degenerate acts and emotional support, succeeded to make Dan resume his gaming. Tina stopped the cattle-scalping acts in Farmtasia, relieved Dan’s anger, and encouraged him not to give up. However, in Bert’s case, unlike Cara, he did not “react” to Tina’s “two-mark intervention.” The opportunity of getting a few more marks contributing to the continuous assessment performance was not appealing to Bert.

Implications

Most of the contemporary COGBlE research has treated students’ game-based learning process as a “black box,” focusing on studying whether the particular game-based learning approaches could yield the particular learning outcomes via positivist experiments. Ferdig (2005) argued that technological educators should not only provide students with pedagogically sound technology, but also learn from students who can offer insights into the ways of making technology more pedagogically sound. The pragmatic integration of COGBlE into education should be an iterative course, requiring the input of empirical understandings of how and why students learn or do not learn in the teaching and learning process.

Adopting a qualitative research approach, the present study uncovered the “black box” inside VISOLE, revealing the significance of teacher facilitation in COGBlE upon the setting of formal teaching in school education. In fact, teacher facilitation is not a “new” pedagogical component of game-based learning. For examples, the importance of teacher debriefing was accented when introducing “non-computer” simulation games to education in the 1980s (Cruickshank & Telfer, 1980; van Ments & Hearnden, 1985). Nonetheless, it has received little attention in the era of
“computer” game-based learning (either EIG or GIE) (Peters & Vissers, 2004). The findings of this study also alert the researchers in the field to the assessment issue related to COGBLe. Indeed, assessment has been regarded as one of the vital elements in education (Morris, 1996). Its purpose is not merely for offering examination-oriented students (like Cara) justifications to participate in learning activities, but more importantly, providing students with feedback to improve or enhance themselves in the process of learning (Biggs & Moore, 1993). However, there has been little in-depth discussion in the literature about how teachers should assess students when COGBLe is harnessed in formal education.

As witnessed in this study, accurately assessing students’ learning performance in COGBLe might not be easy. A game usually has exploits (Salen, 2008); players’ good gaming results might be due to their adoption of degenerate strategies. Thus, relying solely on students’ gaming results to assess their learning outcomes may provoke the issue of assessment unfairness. Evaluating students’ learning performance via reviewing all their gaming proceedings seems a more precise method; however, it will be too time-consuming in practice. Using students’ off-the-game artifacts such as reflective journals as an assessment basis sounds more pragmatic, but it may disadvantage the ones who are bad in language or hate writing. Further effort from the field on looking into the assessment matter in COGBLe upon the context of formal education is definitely necessary.

Limitations

We adopted a single-case study approach in the present study. It is undeniable that the generalizability of our findings is limited. Nevertheless, generalization is not necessarily an aim of every research (Denzin, 1983). One selects this research approach because he/she wishes to understand the particular in-depth, instead of generally true of many (Creswell, 2008). Controlling the teacher-factor, we focused on gaining an in-depth understanding of the impeding phenomena emerging from the course of VISOLE (a particular approach to integrating COGBLe into formal teaching).

Although we can justify our research design, our findings do have limitations. Firstly, we cannot claim that the discovered impeding phenomena will emerge every time when VISOLE is implemented. We conducted this research in a senior secondary Band-2 classroom. The same or new impeding phenomena may or may not emerge in a Band-1 classroom (where students may be more examination-oriented), or a Band-3 classroom (where may have less non-gamers and more gamers), or a primary and junior-secondary classroom (where the examination pressure on students is lower). Secondly, the desirable outcomes of the teacher facilitation that we witnessed are “teacher-dependent.” The teacher participant (Tina) was a VISOLE-competent teacher and volitional to harness VISOLE in her teaching practice. Other teachers (especially non-gamer teachers) may not have the same enthusiasm or competence. Therefore, we cannot claim that any teachers taking the same teacher facilitation acts will result in the same desirable outcomes as evidenced in this research.

Conclusion and future work

This research was an investigation on implementing VISOLE in the context of formal teaching at school. Specifically, we focused on probing into the phenomena, which emerged in the course of VISOLE, impeded students’ learning process. Four students—a non-gamer student, a gamer student, an examination-oriented student, as well as an angry student were our focal units of analysis. In their learning process, we identified three impeding phenomena—impromptu gaming, arbitrary gaming, and halt gaming, and uncovered why and how these phenomena emerged in and impeded their learning. The findings signified the importance of teacher facilitation in COGBLe and gave a new understanding of adopting COGBLe in formal teaching from the students’ point of view. We hope this piece of work can provide an empirical reference for EIG/GIE researchers and education practitioners (such as education policy makers, educators, school teachers, educational game designers, etc.) who are developing, adopting or appropriating games or gaming technology to support teaching and learning.

We should never assume every student (even if he/she is young) is technically-literate in gaming. In our previous design of VISOLE, the scaffolding work (Phase 1) emphasized only on equipping students with high-level abstract knowledge, without paying attention from the technical perspective to preparing them to start up their gaming. We are now designing a number of new NPCs (non-player characters) functioning as virtual game tutors in Farmtasia.
Hence, students can get just-in-time support and/or guidance through interacting with the NPCs at the beginning of Phase 2 of VISOLE. Corresponding evaluation on the effectiveness of these virtual game tutors will be carried out in the near future.

In light of the limitations of the present research, “implementation of VISOLE in a Band-1 classroom and a Band-3 classroom” has been one of the items on our further study agenda. We will also place additional effort on studying how to assess students’ learning performance in the course of VISOLE (and COGBLe in general). Team-based collaboration (Jong, 2014), which is currently absent in the game-play of Farmtasia, is regarded as one of significant strategies to further promote transfer of learning in the process of gaming. We are now exploring the possibility of infusing intra-group and inter-group collaborative elements into Farmtasia so that students can learn not only competitively but also collaboratively during Phase 2 of VISOLE.

Acknowledgments

The work described in this paper was substantially supported by a grant from the Research Grants Council of Hong Kong Special Administration Region, China (Project No.: 459013). We also very much appreciate the scholarly and technical advice provided by Prof. Jimmy Lee, Dr. Lucy Huang, Mr. Eric Luk, and Mr. Peter Chan from The Chinese University of Hong Kong, and Dr. Vincent Tam from The University of Hong Kong.

References


A Learning Style-based Grouping Collaborative Learning Approach to Improve EFL Students’ Performance in English Courses

Yu-Chen Kuo*, Hui-Chun Chu and Chi-Hao Huang

Department of Computer Science and Information Management, Soochow University, 56, Sec. 1, Kui-Yang St., Taipei 100, Taiwan // yckuo@csim.scu.edu.tw // carolhcchu@gmail.com // erica.huang0716@gmail.com

*Corresponding author

(Submitted March 14, 2014; Revised June 30, 2014; Accepted July 25, 2014)

ABSTRACT
Learning English is an important and challenging task for English as Foreign Language (EFL) students. Educators had indicated that, without proper learning support, most EFL students might feel frustrated while learning English, which could significantly affect their learning performance. In the past research, learning usually utilized grouping, but few studies have considered the difference in group members’ learning styles. In this study, a learning style-based collaborative learning approach is proposed to cope with this problem. To evaluate the effectiveness of the proposed approach, an learning style-based online collaborative learning platform has been developed. Moreover, an experiment has been conducted on a university English course to compare the learning performance and learning interest of the students who learn with the proposed approach and those who learn with the conventional collaborative learning approach. It was found that the homogeneous learning style groups outperformed the heterogeneous groups. Besides, those students who used the online English collaborative learning approach outperformed those who used the traditional paper-based English collaborative learning approach. Sequence analysis was also carried out to analyze the students’ online discussion.

Keywords
Architectures for educational technology systems, Improving classroom teaching, Cooperative/collaborative learning, Applications in subject areas, Evaluation of CAL systems

Introduction

Research background and motivation

In the modern environment in which the global village is emphasized, English has become an essential second language as well as a language for international communication. Many researchers have emphasized the importance of English learning (Hsu, Hwang, & Chang, 2013). In most universities in Taiwan, the students must satisfy the English requirements before they can graduate. The TOEIC test (published by Educational Testing Service) is the standardized test for English proficiency and is often required for young people entering a business company as a working ability in Taiwan (Yang, Chuang, Li, & Tseng, 2013). Therefore, getting higher grades of TOEIC test is a common target of undergraduate students nowadays.

With the rapid progress of information technology, many researchers have attempted to apply various online instruction strategies to provide students more efficient way to learn second language via Internet (Chen, Wang, & Chen, 2014; Hsu et al., 2013; Liu et al., 2011; Wen, Looi, & Chen, 2012). Scholars have also indicated that the importance of managing students’ learning conditions, addressing their weaknesses, and leading collaborative learning activities (Chu, Hwang, & Tsai, 2010; Hwang, Chu, Shih, Huang, & Tsai, 2010; Zhan, Xu, & Ye, 2011). Researchers have therefore carried out relevant research on students interacting and learning online with their peers. It can be seen that learning by sharing information and exchanging knowledge among peers could not only enhance the competitiveness of students, but could also promote their learning performance (Cooper & Cowie, 2010; Hwang, Chu, Lin, & Tsai, 2011; Wang, 2010; Wen et al., 2012).

As a way of ensuring students’ online learning interaction and discussion, the grouping method has been discussed widely based on students’ characteristics (AbuSeileek, 2012; Chiu & Hsiao, 2010; Moreno, Ovalle, & Vicari, 2012) or their learning achievements (Lin, Huang, & Cheng, 2010; Wang, 2010). Grouping students for collaborative learning allows them to share their ideas and learning experiences, and further promotes the learning performance of the group as well as of the individuals (Huang, Wu, & Chen, 2012; Wang, & Hwang, 2012). Lin, Huang, and Cheng
considered “student level” and “interests” as two important criteria for grouping students. Hsieh et al. (2011) indicated that students could be benefited more if they received the teaching strategies matched their learning style were taken into account in learning design.

Kolb (1984) classified learning styles into Diverger, Assimilator, Converger, and Accommodator. With different learning methods, distinct learning styles could provide adaptive learning for enhancing learning performance (Mampadi, Chen, Ghinea, & Chen, 2011; Tseng, Chu, Hwang, & Tsai, 2008). Adán-Coello et al. (2011) further indicated that group members’ learning styles might affect peer interactions in collaborative learning activities. Therefore, this study investigates the effects of different learning style-based grouping strategies on collaborative learning environments in order to compare the interaction among students who were grouped with the same and different learning styles, and to investigate the effects of the two different grouping strategies on the online collaborative learning environment for each of Kolb’s learning styles. For instance, Diverger students presenting the same information processing skills might suit collaborative learning in a group with students of the same learning style.

Moreover, in this study, to further understand the students’ interaction patterns in the asynchronous discussion of the proposed English collaborative learning platform, the students’ discussion content has also been analyzed. The interaction patterns of students in groups with different learning styles are further explored. The effects of English collaborative learning with a focus on TOEIC are investigated. Meanwhile, the effects of learning style-based grouping on English collaborative learning are discussed.

Research purposes

The research method involves collaborative learning and introduces learning styles to group students with the same learning styles into homogeneous groups, and others with different learning styles into heterogeneous groups. The learning performance and behavior of the students who learned with different grouping strategies (homogeneous, heterogeneous, or random) in the English course were analyzed to accomplish the following research purposes:

• To investigate whether the online collaborative learning platform could enhance students’ English learning performance.
• To compare the learning performance of the students who learned with different grouping strategies (homogeneous group, heterogeneous group, and random) in the English course.
• To find a suitable way of using Kolb’s Learning Styles for grouping students.
• To compare the interaction patterns of the students assigned to the collaborative learning groups based on the learning style-based homogeneous and heterogeneous grouping strategies.

Related work

Online collaborative learning and English learning

The convenience of the Internet has changed the traditional face-to-face collaborative learning; that is, students can cooperatively complete their learning tasks online. Many researchers have therefore invested strategies of online collaborative learning. For instance, Wang (2010) proposed the application of an online sharing platform for collaborative learning. With the aim of increasing the effectiveness of team teaching, Cooper and Cowie (2010) mentioned that collaborative learning allows students to mutually create new knowledge. Recently, online collaborative learning has become a popular research issue.

English is a language for communication. During independent learning, students cannot effectively promote their learning performance because of the lack of interaction and discussion practices. For this reason, many schools in Taiwan have opened TOEIC courses for students to learn together in class. By in-class learning, students could interact with the teachers, mutually learn among peers, and solve problems on the spot. Nevertheless, the time and location for in-class learning are restricted. Due to the popularity of the Internet, researchers have proposed the benefits of collaborative learning for students’ English reading capability in TOEIC (Yang, Chuang, Li, & Tseng,
However, few researchers have discussed the effects of learning styles on students’ English collaborative learning. This study therefore explores this aspect in depth.

**Kolb learning style**

Kolb (1984) classified learning styles into Diverger, Assimilator, Converger, and Accommodator as shown in Figure 1. Diversers (Feeling and Watching, CE/RO) tends to watch from distinct aspects, proposes specific opinions and ideas, prefers watching to doing, enjoys working with others, and listens to others’ opinions with an open mind. Accommodators (Doing and Feeling, CE/AE) prefers to prove things with actions, is likely to be attracted by new ideas and challenges, does things without logical thinking, shows courage and insistence, and enjoys completing work with group cooperation. Convergers (Doing and Thinking, AC/AE) shows excellent ability to look for suitable solutions, prefers challenging technical tasks, but lacks interest in people or interpersonal interaction. They are good at doing and are brave in the face of problems and reveal better ability to solve technical problems than to exchange ideas or to have interpersonal interaction. Assimilators (Watching and Thinking, AC/RO) prefers simple and logical thinking, focuses more on ideas and concepts than on others’ opinions, requires clear and powerful explanations, presenting them with clear logic, excels at logic, theories, and reasoning, and favors reading, exploring, and thinking.

![Figure 1. Kolb’s (1984) Learning Styles](image_url)

Many researchers have discovered the effects of learning styles on learning performance (Hwang, Sung, Hung, & Huang, 2013; Mampadi et al., 2011; Tseng et al, 2008). For example, both Diversers and Accommodators present feeling, suggesting that they might favour group collaborative learning. In this case, students with such learning styles being grouped together to form a homogeneous group might exhibit better online collaborative learning performance than students with other learning styles. Huang et al. (2011) proposed heterogeneous grouping for students to collaboratively learn the engineering curriculum and showed good learning performance of the heterogeneous group. Adán-Coello et al. (2011) proposed homogeneous and non-homogeneous grouping for students to collaboratively learn a programming language and showed that the homogeneous group outperformed the other groups, and argued that the homogeneous group, because it is made up of students with the same learning style, can avoid undesirable conflict and can easily achieve a consensus solution. Consequently, how to group students become an important issue for leading a collaborative learning activity.

Previous studies mainly focused on investigating the effects of technology-enhanced learning approaches on learning performances of the students with different learning styles (Chu, 2014; Huang et al., 2011; Hwang, Sung, Hung,
Huang, & Tsai, 2012; Zhan et al., 2011). For example, Hwang, Sung, Hung, Huang and Tsai (2012) developed an adaptive educational computer game by taking students’ learning styles into consideration and found that playing style-fit games significantly improved the students’ learning achievements. Only few studies have attempted to investigate the issues related to learning style-based grouping in collaborative learning; for example, Kyprianidou et al. (2012) compared the communication frequencies of the students who learned with style-based heterogeneous grouping approach and the random grouping approach. To our knowledge, none of these studies have compared the effectiveness of learning style-based heterogeneous and homogeneous grouping strategies from various dimensions of learning performances and learning behaviors. Therefore, the purpose of this study is to develop a learning style-based collaborative learning platform and explore the collaborative learning behaviors and performances of the students who learn in homogeneous and heterogeneous groups formed based on the four dimensions of Kolb’s Learning Styles (Feeling, Doing, Thinking, & Watching).

Research method

The design of collaborative learning strategy and learning environments

An online collaborative learning platform with TOEIC English tests as the learning content was established. Students collaboratively interacted with their peers in groups of homogeneous or heterogeneous learning styles. With gradual guidance, learners who spend sufficient time on a learning process will transfer the learning results into knowledge. Furthermore, learners could deliver their knowledge to their peers and create new knowledge by collaborative interaction and discussion. The collaborative learning strategy in our online learning system is shown in Figure 2. Students log into the system (Step 1) and then start to answer the questions (Step 2). When a student completes a question, he/she has to wait for other group members to complete the question (Step 3). After all members have completed the question, the system detects the consistency of their answers (Step 4). When there are different answers, the group members have to discuss (Step 4.1) and revise their answers (Step 2) until a consistent answer is provided (Step 4). The consistent answer is then verified (Step 5). When the answer is correct, the next question is shown (Step 6). If the answer is incorrect, the system will check whether it is the first incorrect answer or not (Step 5). If it is the first incorrect answer, the group members have to discuss (Step 4.1) the question again (Step 2). If there is a second incorrect answer, the system skips to the next question so as to prevent the students from repeatedly guessing the answer.

Figure 2. Online collaborative learning strategy
The students in the control group were grouped randomly and used the conventional face-to-face paper-based collaborative learning strategy as shown in Figure 3. The test paper is distributed to each group of students (Step 1) for discussion (Step 2). In the process, the students share and discuss the questions to come up with a common answer depending only on the English knowledge of the group members (Step 3). They continue to the next question (Step 4) until the test paper is handed in (Step 5).

![Figure 3. Face-to-face paper-based collaborative learning strategy](image)

**The interface design of online learning system**

The home page of the online collaborative learning platform system was designed using PHP and Dreamweaver. After logging in, the system displays the test frame as shown in Figure 4, where the functions of the online collaborative learning platform are introduced. The frame is divided into several areas as shown below.

![Figure 4. System contents page of the online collaborative learning platform](image)
Block 1 is the question area, displaying the TOEIC question. After answering the question, the students can select “next” for the system to verify the correctness of their answers. If their answers are correct, the next question is shown. Block 2 displays the answering conditions of the group members, where they can be viewed by other students. The system verifies the correctness of the answers only when the answers of the group members are consistent; otherwise, the group members are required to discuss and revise their answers so as to come up with a consistent answer.

The students directly click on the answer in Block 3 to instantaneously update the answer in Block 2. Block 4 provides the discussion area for the group, where the dialogues and the time for answering the question are recorded. In Block 5, the students can choose some common message to quickly respond with comments such as “I don’t know the answer” and “Please answer the question quickly,” to their peers. Block 6 is the message area, where the students type personal messages for collaborative learning. Block 7 shows the list of the online group members.

**Experiment design**

The first independent variable is the learning-based grouping strategy (homogeneous, heterogeneous or random). The second independent variable is the collaborative mode (online or face-to-face). The dependent variable is the students’ learning achievement, while the control variables are the instructor, the learning content, and the prompts provided to the students.

**The participants and experiment procedure**

A total of 48 graduate students aged 23~25 in a university were selected as the experimental subjects. There were 12 Diversers (hereinafter referred to as Div.), 12 Accommodators (Acc.), 12 Convergers (Con.), and 14 Assimilators (Ass.) grouped into 4 homogeneous groups of the same learning style (Experimental group A), 4 heterogeneous groups with different learning styles (Experimental group B), and 4 random groups which did not consider learning style (the Control group), as shown in Figure 5.

*Figure 5. The grouping method of the three groups*
The experiment procedure for the collaborative learning activity is shown in Figure 6. These three groups then took the pre-test for 20 minutes. Experimental groups A and B further utilized the online English collaborative learning platform three times, with 10 English questions practiced each time, for group discussion and learning. The learning period for these groups was not restricted so as to assess the effects of learning period on learning performance. The students could spend almost 120 minutes to finish ten questions. After completing the learning activities, the experimental post-test was given for 20 minutes, aiming to test the enhancement in the students’ English learning achievement.

![Figure 6. Experiment design](image)

**Research tools**

The English test questions were acquired from the book of “New TOEIC Grammar Tests” written by Chin, published by International Books (Chin, 2012). 32 pre-test questions were randomly selected from the book focusing on English grammar to understand the students’ level of English for TOEIC. Another 32 post-test questions were also randomly selected from the book. The aim was to understand the effects of learning style-based grouping collaborative learning on students’ English grammar and vocabulary. Students in the two Experimental groups as well as in the Control group used the same English test questions as their learning material.

Two analyses in non-parametric tests were utilized for the experimental data analyses. The Kruskal-Wallis one-way analysis of variance by ranks (H test) is suitable for the comparison of more than three groups with fewer test samples which applied ranks to analysis of variance. The Mann-Whitney test (U test) is suitable for two independent samples, and aims to test two independent groups with the same parameters. Especially, research with fewer samples would apply such an analysis instead of the t-test for testing the differences between two groups (Conover, 1989).

After completing the learning activities, the chat room dialogues were processed for message coding and sequential analysis, using the message coding proposed by Hou, Sung, and Chang (2009) to analyze the students’ behaviors and attitudes in the chat rooms. The different kinds of behavior and an example of each are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Message coding in chat rooms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Behavior</strong></td>
</tr>
<tr>
<td>C1 Peers knowing each other</td>
</tr>
<tr>
<td>C2 Discussion and comparison</td>
</tr>
</tbody>
</table>
Results

Prior knowledge of pre-test

Before the learning activities, the pre-test was given to test the differences in the basic English knowledge of the 48 students. The Kruskal-Wallis (H test) analysis shows that the mean and the standard deviation for Experimental group A were 42.75 and 25.41 (the lower level presenting the lower scores); for Experimental group B they were 38.82 and 23, and for the Control group they were 42.59 and 25.09, respectively. The significance level \( p = 0.868 > 0.05 \) showed that the scores of these three groups did not reach significant difference, indicating that the students’ basic English knowledge was similar prior to the experiment.

According to the H test, the mean and the standard deviation for Diversers were 33.75 and 5.75, for Accommodators they were 34.5 and 4.88, for Convergers they were 47.25 and 11.50, and for Assimilators, they were 46.5 and 11.88, respectively. The significance level \( p = 0.062 > 0.05 \) indicated no significant difference, meaning that the basic English knowledge of students of the four different learning styles in Experimental group A were similar.

The same process was performed for Experimental group B and it was found with the H test that the mean and the standard deviation for Diversers were 45 and 9.38, for Accommodators they were 45 and 9.38, for Convergers, 41.25 and 8.25, and for Assimilators, they were 38.25 and 7.00, respectively. The significance level \( p = 0.877 > 0.05 \) revealed that the four groups did not reach significant difference; thus, the students’ basic English knowledge in Experimental group B was also similar.

The post-tests of three groups of students

Based on the Kruskal-Wallis (H test) analysis (in Table 2), the significance level \( p = 0.000 < 0.05 \) presented significant difference among these three groups. Post hoc analysis showed that Experimental groups A and B performed better than the Control group, providing evidence that the online English collaborative learning platform could assist the students in terms of their learning performance.

Utilizing the Mann-Whitney (U test) for pair comparison, the mean and standard deviation for Experimental group A were 73.63 and 14.80, and for Experimental group B they were 65.38 and 14.17, respectively. The significance level \( p = 0.196 > 0.05 \) presented no significant difference in the learning performance of these two groups, both of which used the online collaborative learning platform.

<table>
<thead>
<tr>
<th>( p )</th>
<th>Post hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>(1)&gt;(3)</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

The post-test results of homogeneous and heterogeneous groups

The learning performance of the homogeneous and heterogeneous groups and the suitability of the learning styles for such grouping are discussed in this section. According to the Kruskal-Wallis (H test) analysis, Accommodators and Convergers revealed better scores than Diversers and Assimilators in the homogeneous groups in Experimental group A, as shown in Table 3. From the Post hoc analysis, the significance level \( p = 0.009 < 0.05 \) shows that
Accommodators and Convergers outperformed Divergers and Assimilators; that is, the students with different Kolb’s Learning Styles exhibited significant differences in their learning performance in the homogeneous groups.

| Table 3. H test analysis of the post-test in the homogeneous groups (Experimental group A) |
|---|---|---|---|---|---|---|
| Homogeneous group | N | Mean | SD | Mean Rank | df | p |
| (a) Divers | 4 | 60.25 | 9.74 | 3.88 | 3 | 0.009 |
| (b) Accommodators | 4 | 86.25 | 9.75 | 12.50 | (b)(c)> |
| (c) Convergers | 4 | 85.25 | 6.71 | 12.50 | (a)(d) |
| (d) Assimilators | 4 | 62.75 | 8.46 | 5.13 |

According to the Kruskal-Wallis (H test) analysis, similar scores were revealed for the different learning styles in the heterogeneous groups in Experimental group B, as shown in Table 4. The significance level \( p = 0.501 > 0.05 \) presented no significant difference for Kolb’s Learning Style in the learning performance of the heterogeneous groups.

| Table 4. H test analyses of the post-test in the heterogeneous groups (Experimental group B) |
|---|---|---|---|---|---|---|
| Heterogeneous groups | 1 | 2 | 3 | 4 | N | Mean | SD | Mean Rank | df | p |
| Div. Div. | 4 | 69.50 | 20.42 | 9.00 | 3 | 0.501 |
| Acc. Acc. | 4 | 68.75 | 15.43 | 5.50 |
| Con. Con. | 4 | 68.50 | 10.12 | 10.38 |
| Ass. Ass. | 4 | 54.75 | 6.65 | 9.13 |

Diverger (Div.), Accommodator (Acc.), Converger (Con.), and Assimilator (Ass.)

From the experimental data, the mean and the standard deviation of Divergers in the homogeneous group were 60.25 and 9.74, while they were 69.5 and 20.42 in the heterogeneous group, respectively. The significance level \( p = 0.686 > 0.05 \) showed no significant difference in the learning performance of Divergers in these two types of groups. Nevertheless, Divergers’ scores in the heterogeneous group were higher than those in the homogeneous group. The mean and the standard deviation of Accommodators in the homogeneous group were 86.25 and 9.75, and 68.75 and 15.43 in the heterogeneous group, respectively. The significance level \( p = 0.200 > 0.05 \) revealed no significant difference in the learning performance of Accommodators in these two types of groups. However, the Accommodators’ scores in the homogeneous group were higher than those in the heterogeneous group.

The mean and the standard deviation of Convergers in the homogeneous group were 85.25 and 6.71, and 68.50 and 10.12 in the heterogeneous group, respectively. The significance level \( p = 0.057 > 0.05 \) presented no significant difference in the learning performance of Convergers in these two types of groups. However, the scores between the homogeneous and heterogeneous groups showed differences in that the Convergers’ scores in the homogeneous group were higher than those in the heterogeneous group. The mean and the standard deviation of Assimilators in the homogeneous group were 62.75 and 8.46, and 54.75 and 6.65 in the heterogeneous group, respectively. The significance level \( p = 0.486 > 0.05 \) revealed no significant difference in the learning performance of Assimilators in these two types of groups. Nonetheless, Assimilators’ scores in the homogeneous group were higher than those in the heterogeneous group.

Rather than having students cooperatively complete learning tasks, in this study students had to answer the same question based on consistent ideas and concepts so that the effect of drawing on the strong points of the different learning styles would not appear. Relatively, students with different learning styles collaboratively learning and negotiating for consistent answers would not necessarily present the effect of drawing on their strong points. On the other hand, students with the same learning style could easily communicate, negotiate, and achieve a consistent answer. Thus, as the results show in Table 2, although the learning performance of the homogeneous group was not significantly different from that of the heterogeneous group, the mean of the post-test in the homogeneous group outperformed that of the heterogeneous group.
The time period of learning process

In this study, the average learning period was 20–25 minutes for Experimental group A, 35–40 minutes for Experimental group B, and 25–30 minutes for the Control group. It was found that Experimental group A (homogeneous group) had a shorter learning period; therefore, a consistent answer could be rapidly achieved during the collaborative learning. Regarding the Control group with the paper-based collaborative learning, they needed to refer to the prompts from the paper-based reference book; therefore, they repeatedly answered the question when they gave an incorrect answer, so the students’ learning period in the control group was longer than that of Experimental group A. Experimental group B (heterogeneous group) had the longest learning period, meaning that they spent more time on the collaborative discussion in order to achieve a consistent answer. In contrast, we found that students with the same learning style could easily communicate, negotiate, and rapidly achieve a consistent answer.

Sequential analysis

After the students finished using the online learning platform, the communication messages from the chat rooms were coded. Those messages related to self-introduction or knowing each other were coded as C1 (“Peers knowing each other”), messages related to discussion of the question or comparison of members’ ideas were coded as C2 (“Having discussion”), messages proposing solutions were coded as C3 (“Proposing solutions”), messages seeking the answer were coded as C4 (“Looking for solutions”), and messages containing unrelated responses were coded as C5 (“Irrelevent response”). Scorer Reliability Analysis of the coding revealed that the Spearman-Brown Coefficient = 0.937. The proportions of message codes in the homogeneous group are shown in Table 5. From the proportion data, Accommodators and Convergers produced more messages, about 34.87(%) and 33.72(%), respectively, of the total number of messages in the homogeneous group, and among these two types of groups, C3 (Proposing solutions) and C4 (Looking for solutions) presented high proportions. Consequently, Accommodators and Convergers seem to be more suitable for learning in homogeneous groups. The proportions of message codes in the heterogeneous group are listed in Table 6. The proportions of the codes among the various groups were close, and C5 (Irrelevant response) appeared with the highest proportion, suggesting that more communication and negotiation were required for the heterogeneous group, thus influencing their performance. The results are consistent with the conclusions in Table 2. Table 3 and Table 4 in that the mean of the post-test in the homogeneous group outperformed that of the heterogeneous group. Accommodators and Convergers presented better learning performance in the homogeneous group, whereas they showed no significant difference for the learning performance in the heterogeneous group.

<table>
<thead>
<tr>
<th>Homogeneous</th>
<th>C1 (%)</th>
<th>C2 (%)</th>
<th>C3 (%)</th>
<th>C4 (%)</th>
<th>C5 (%)</th>
<th>Number of message (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Div.</td>
<td>0.69</td>
<td>0.46</td>
<td>5.31</td>
<td>0.92</td>
<td>11.09</td>
<td>18.48</td>
</tr>
<tr>
<td>Acc.</td>
<td>0.92</td>
<td>1.62</td>
<td>11.09</td>
<td>9.70</td>
<td>11.55</td>
<td>34.87</td>
</tr>
<tr>
<td>Con.</td>
<td>0.92</td>
<td>1.62</td>
<td>12.24</td>
<td>7.85</td>
<td>11.09</td>
<td>33.72</td>
</tr>
<tr>
<td>Ass.</td>
<td>0.46</td>
<td>1.39</td>
<td>3.00</td>
<td>0.46</td>
<td>7.62</td>
<td>12.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Heterogeneous</th>
<th>C1 (%)</th>
<th>C2 (%)</th>
<th>C3 (%)</th>
<th>C4 (%)</th>
<th>C5 (%)</th>
<th>Number of message (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>0.71</td>
<td>1.95</td>
<td>4.80</td>
<td>2.84</td>
<td>15.63</td>
<td>25.93</td>
</tr>
<tr>
<td>Group II</td>
<td>0.36</td>
<td>1.07</td>
<td>2.84</td>
<td>1.07</td>
<td>14.56</td>
<td>19.89</td>
</tr>
<tr>
<td>Group III</td>
<td>0.53</td>
<td>3.20</td>
<td>7.99</td>
<td>2.13</td>
<td>14.39</td>
<td>28.24</td>
</tr>
<tr>
<td>Group IV</td>
<td>0.36</td>
<td>4.44</td>
<td>5.86</td>
<td>2.84</td>
<td>12.43</td>
<td>25.93</td>
</tr>
</tbody>
</table>

The message codes were further analyzed using the Sequential Analysis approach proposed by Hou, Sung, and Chang (2009). Sequential Analysis describes the message code sequence with the significance (Z) of the Z fraction binomial
test to present the learning behaviors with a message code transfer diagram. Each circle in the diagram represents a message code. Each thicker line represents the level of the significance where the significance $Z > 1.96$ and the arrow points in the direction of the transfer. The learning behaviors of the homogeneous and heterogeneous groups are described below in terms of their message code transfer diagrams.

From the message code transfer diagram, as shown in Figure 7, Diverger students in the homogeneous group continuously focused on C3 (“Proposing solutions”) with the significance $Z = 5.88$. Diverger students in the homogeneous group did indeed often propose their own opinions and ideas. For example, user_1 said, “I think the answer is variety.” Sometimes, C3 (“Proposing solutions”) would be transferred to C5 (“Irrelevant response”) with the significance $Z = 4$. For example, user_3 said, “Hurry up!” However, the significance of transferring from C5 (“Irrelevant response”) to C3 (“Proposing solutions”) reached $Z = 2.74$. Since Divergers tended to propose their own opinions, the students would transition between C3 (“Proposing solutions”) and C5 (“Irrelevant response”) before coming up with a consistent answer.

As shown in Figure 7, Assimilator students in the homogeneous group showed the significance $Z = 6.08$ for C5 (“Irrelevant response”) and it appears that they spent more time thinking. Moreover, Assimilator students in the homogeneous group presented the significance $Z = 2.81$ on the transition from C2 (“Having discussions”) to C3 (“Proposing solutions”), revealing a two-way connection line. Apparently, Assimilator students in the homogeneous group would not only think independently, but would also occasionally discuss with others. For example, user_8 said, “Wrong! The prompt shows…” Finally, both the Diverger and Assimilator groups showed a high occurrence of C5 (“Irrelevant response”) to the extent that their learning performance was affected.

![Message code transfer diagrams for the homogeneous Diverger and Assimilator groups](image)

Accommodators tend to listen to their peers’ opinions while learning. As shown in Figure 8, Accommodators in the homogeneous group showed the significance $Z = 2.2$ for transferring from C2 (“Having discussions”) to C3 (“Proposing solutions”). In this case, the Accommodator group proposed dialogues related to correct answers after discussion with their peers. For example, user_16 said, “The question should be… so the answer is B!” Accommodator students in the homogeneous group therefore tended to listen to their peers’ ideas. Besides, Accommodator students in the homogeneous group would transfer from C3 (“Proposing solutions”) to C5 (“Irrelevant response”) with the significance $Z = 2.2$. For instance, user_14 said, “I will follow your answers.” However, the students would transfer from C5 (“Irrelevant response”) back to C3 (“Proposing solutions”) with the significance $Z = 2.78$. Since Accommodators tended to listen to their peers’ ideas and opinions, the dialogues transferred between C3 (“Proposing solutions”) and C5 (“Irrelevant response”) before the peers’ ideas and opinions were expressed.

Moreover, as shown in Figure 8, Convergers in the homogeneous groups would continuously repeat C3 (“Proposing solutions”) and C4 (“Looking for solutions”) statements with the significance $Z = 4.66$ and 3.34, respectively. It appeared as a one-way connection line with the significance $Z = 2.02$. Consequently, Converger students in the homogeneous groups were actually brave in the face of problems and did directly propose questions for discussion with their peers. For example, user_9 said, “Is the answer D?” Moreover, Converger students in the homogeneous
group would transition from C3 (“Proposing solutions”) to C5 (“Irrelevant response”) with the significance $Z = 4.22$. They would deviate from the topic when waiting for others’ answers, but would return to the topic quickly. For instance, user_10 said, “Hurry up.” They therefore transferred between C3 (“Proposing solutions”) and C5 (“Irrelevant response”). Finally, both the Accommodator and the Converger groups presented the characteristic of Doing, meaning that the transition among various statuses appeared on the message transfer chart, which presented the active interaction among peers and fewer occurrences of C5 (“Irrelevant response”). From the sequential analysis results for the homogeneous group, we observe that the learning performance of the Accommodator and Converger groups were better than that of the Diverger and Assimilator groups, which was consistent with the observation shown in Table 3.

![Figure 8. Message code transfer diagrams for the homogeneous Accommodator and Converger groups](image)

Regarding the heterogeneous groups, students with different learning styles learned together in each heterogeneous group (as shown in Figure 9). The learning behaviors in the heterogeneous groups mostly transitioned from C3 (“Proposing solutions”) to C5 (“Irrelevant response”). For instance, user_21 said, “The answer is B”, but then user_24 said, “I don’t know!” However, most students in the heterogeneous groups would transition from C5 (“Irrelevant response”) back to C3 (“Proposing solutions”), forming a two-way connection line. These four heterogeneous groups showed high occurrences of C5 (“Irrelevant response”) with the significance $Z = 19.94, 30.15, 10.49,$ and $15.84$, respectively. Thus, students in the heterogeneous groups were likely to present their own opinions because of their different learning styles. The discussion period was therefore prolonged as it was difficult to reach a consensus. Besides, the heterogeneous groups’ occurrence of C5 (“Irrelevant response”) was higher than that of the homogeneous groups. Consequently, the learning performance of the heterogeneous groups showed no significant difference, which is consistent with the conclusion shown in Table 4.

![Figure 9. Message code transfer diagram for the heterogeneous groups](image)
Discussion and conclusions

Research on learning styles has been greatly emphasized, as students’ learning styles affect their learning intentions. Thus, this study implemented the English collaborative learning system to investigate the effects of learning style-based grouping. The collaborative learning method adopted in this study required the students to solve the same question with only one answer which they should arrive at through discussion and negotiation in order to reach a consistent answer. Due to this approach, the effect of drawing on the strong points of students with different learning styles would therefore not necessarily appear as they had to negotiate a consistent answer.

In terms of the suitable way of using Kolb’s learning style for grouping students, there was no significant difference in the learning performance of Experimental groups A and B. However, the students in Experimental group A with homogeneous learning styles had more consistent opinions, so the time required to complete the learning activities was shorter than that of Experimental group B, the groups made up of students with heterogeneous learning styles. Furthermore, the learning performance of Accommodators and Convergers when working together in homogeneous groups was better, as these two kinds of learners tend to positively enjoy interacting and discussing problems with others, according to Kolb’s Learning Styles. In Experimental group B, the learning performance in the heterogeneous group does not reveal notable differences. It is speculated that students with different learning styles show different learning attitudes and spend a lot of time on communication and negotiation, with the result that more occurrences of C5 (“Irrelevant Response”) affected their learning performance.

This study aimed to compare the interaction patterns of the students assigned to collaborative learning groups based on the two learning style-based grouping strategies. The sequential analysis showed that both Convergers and Accommodators in the homogeneous groups would properly transfer to “Having Discussions” or “Looking for solutions” and actively interact with their peers; thus, they are suitable for working together in a homogeneous group. Moreover, Diversers and Assimilators produced more instances of “Irrelevant Response” so that their learning performance was less impressive. In the heterogeneous group, the students are likely to hold their own opinions because of different learning methods and do not easily reach a consensus, so that their learning performance was less significant. Besides, the occurrences of “Irrelevant Response” were higher than for the homogeneous group. As a result, providing learners with a suitable learning environment and proper learning methods would assist them in enhancing their learning performance and interest.

To sum up, the findings of this study are helpful to those researchers who try to develop grouping method based on students’ learning style or conduct English online learning activities in the future. Although the learning performance of the homogeneous groups was not significantly different from that of the heterogeneous groups, the mean of the post-test in the homogeneous groups outperformed that of the heterogeneous groups. This is speculated that students with the same learning style could easily negotiate to achieve a consistent answer, and thus improve their learning performance. Our finding is consistent with the observation that students with the same learning style can avoid undesirable conflict and can easily achieve a consensus solution (Adán-Coello et al., 2011). It should be noted that arriving at the right answer more efficiently or quickly does not necessarily mean that greater learning took place; therefore, further studies are needed to examine whether deeper learning have occurred by employing some measuring tools of higher order thinking.

On the other hand, it has been widely debated as to whether learning styles exist, or have any significant impact on learning, implying that further studies and discussions are needed to examine the effectiveness of the learning style-based grouping strategy. The authors of this study are glad to provide the data of the experiment upon requests.

Acknowledgements

This study is supported in part by the National Science Council of the Republic of China under contract number NSC 102-2221-E-031 -002 and NSC 102-2511-S-031 -001.
References


Learning Mathematics with Interactive Whiteboards and Computer-Based Graphing Utility

Ayhan Kursat Erbas1*, Muge Ince1 and Sukru Kaya2

1Department of Secondary Science and Mathematics Education, Middle East Technical University, Ankara, Turkey // 2The Scientific and Technological Research Council of Turkey, Ankara, Turkey // erbas@metu.edu.tr // esdef2001@yahoo.com // sukru.kaya@tubitak.gov.tr

*Corresponding author

(Submitted March 15, 2014; Revised June 23, 2014; Accepted July 28, 2014)

ABSTRACT

The purpose of this study was to explore the effect of a technology-supported learning environment utilizing an interactive whiteboard (IWB) and NuCalc graphing software compared to a traditional direct instruction-based environment on student achievement in graphs of quadratic functions and attitudes towards mathematics and technology. Sixty-five high school graduates attending cram schools (called dershane in Turkish) to study for the university entrance examination participated in the study. The significant interaction effect between time of testing and groups indicated that student performance across time of testing was dependent upon the presence or absence of the treatment. Comparisons suggest that although both experimental and control group students’ performances increased from pre-test to post-test and then decreased from post-test to retention test, the rate of decrease was about the same and the rate of increase was different, with students showing a greater rate of increase in the experimental group. The results also revealed that the treatment had positively affected students’ attitudes towards technology and mathematics. Moreover, students’ reasoning and interpretation skills regarding graphs of quadratic functions were better in the experimental group as compared to those in the control group.

Keywords

Interactive whiteboard, Computer-assisted learning, Graphing, Quadratic functions, Mathematics education

Introduction

IWBs may turn a computer into a powerful teaching and learning device with the use of vivid colors, different fonts, integrated web pages and applets, presentation software, and scanned images (Fernandez & Luftglass, 2003). In recent years, large-scale investments in IWB technology throughout the world have been quite striking (see Digregorio & Sobel-Lojeski, 2010; Hennessy & London, 2013; Kitchen, Finch, & Sinclair, 2007; Slay, Siebörger, & Hodkinson-Williams, 2008; Somyurek, Atasoy, & Ozdemir, 2009). For example, a recent effort called “Movement of Enhancing Opportunities and Improving Technology,” abbreviated as the “FATIH Project” (Turkish Ministry of National Education [MoNE], 2011), proposes that a “Smart Class” project be put into practice in all schools around Turkey. With this project, 42,000 schools and 570,000 classes will be equipped with the latest information technologies and turned into computerized education classes (Smart Classes) by providing tablets and LCD touch screen interactive boards. As part of the project, each of the around 630,000 teachers will receive laptops. But is it worth it?

Several studies suggest that use of IWBs can have positive effects on teaching and learning (British Educational Communications and Technology Agency [BECTA], 2003) as they promote students’ interest and motivation, provides longer sustained concentration and more effective learning (Glover & Miller, 2001; Somekh et al., 2007; Wood & Ashfield, 2008) and foster responses to a range of learning processes (Glover, Miller, Averis, & Door, 2007). However, much of the literature supporting the use of the IWBs provides insubstantial evidence (Clarkson, 2011; Smith, Higgins, Wall, & Miller, 2005). The literature regarding IWBs in mathematics education, on the other hand, is limited in terms of how they affect students’ attitudes toward technology in mathematics and how their achievement is affected by this new technology (Heemskerk, Kuiper, & Meijer, 2014). The current research emphasizes mostly how to use these new devices (e.g., Miller, Glover, & Averis, 2005; Shae & Tseng, 2001), teacher use of IWBs (e.g., Beauchamp, 2004; Kennewell & Beauchamp, 2007), the pedagogy of using IWBs, student attitudes, and interaction among students, teacher, and technology (Armstrong, Barnes, Sutherland, Curran, Mills, & Thompson, 2005; Wall, Higgins, & Smith, 2005). For example, after investigating if the concepts of rotation, translation, and reflection can be taught by IWB more effectively in seventh grades, Robinson (2004) concluded that the IWB has no significant effect on achievement but improves the attitudes towards the lesson. Similarly,
Heemskerk, Kuiper, and Meijer (2014) found no evidence that the use of an IWB in mathematics lessons is associated with better learning results, while students with higher exposure to IWB lessons are more motivated for mathematics and perform better than students with lower exposure to IWB lessons. So why and how to use IWBs in mathematics education is still a question open to empirical test.

Purpose of the study

As in the case of the FATİH Project in Turkey, widespread technology initiatives and investments worldwide require empirical evidence about if and how the intended technologies would be beneficial to the students and teachers. Without such evidence, while there are less expensive projection technology, justification of IWB use for enhancing students learning beyond being a presentation tool is problematic (Wall, Higgins, & Smith (2005). According to Moss and Carey (2010, pp. 20-23), the effect of IWBs depends crucially on how it is used, not on the mere absence or presence in the classroom. Considering that IWBs are growing popular even without such strong evidence, then, as Ringstaff and Kelly (2002) once asked about education technology in general, the important question should be “Under what conditions do IWBs have the most benefit for students?” rather than “Are IWBs worth the cost?” Moreover, as Digregorio and Sobel-Lojeski (2010) concluded in their extensive review of the research on the effects of IWBs on student performance and learning, the evidence of the effect of IWB use on students’ learning and motivation is inconclusive and research regarding IWBs should focus beyond generic learning gains and on subject-specific learning, along with taking the contextual factors and mediating variables into account. Geiger, Forgasz, Tan, Calder, and Hill (2012) conclude that while IWBs are becoming increasingly available, there is little research identifying “how best to use it to promote deep mathematical knowledge and understanding, or to enhance genuinely collaborative approaches to learning” (p. 132). Regarding the role of IWBs for enhancing the quality of interaction and improving conceptual mathematical understanding, De Vita, Verschaffel, and Elen (2014) concludes that “there is need for greater attention to the pedagogy associated with IWB use and, more specifically, to stimulate the design of new kinds of learning environments.” Thus, the purpose of the present study was to investigate the effects of a technology-supported learning environment utilizing IWBs in conjunction with NuCalc graphing software (Avitzur, Gooding, Wales, & Zadrozny, 1998) on students’ achievement in quadratic equations and functions compared to traditional direct instruction. Furthermore, students’ attitudes towards technology and mathematics and their views about using IWBs and computers in learning mathematics were also studied.

Technology and mathematics education: Case of graphs and graphing

Graphs and graphing are particularly considered by many teachers and researchers to be the most important and fundamental concepts in all of mathematics (Doerr & Zangor, 2000; Leinhardt, Zaslavsky, & Stein, 1990; Mevarech & Kramarsky, 1997), since graphs are essential for school algebra and can be used as a bridge between concrete thinking and abstract thinking (Piaget, Grize, Szeminska, & Bang, 1977). As technological environments support visualization and experimentation (Arcavi & Hadas, 2000), technology has allowed much more emphasis to be placed on graphs and their interpretation. They offer many advantages over manual graph plotting, such as multiple representations, and they also allow students to examine many more graphs more quickly with a high degree of accuracy and with minimal input effort (Forster, 2006; Fung, Hennessy, & Scanlon, 2001; Pierce, Stacey, & Barkatsas, 2007). With technology, students can plot the graphs and then observe their patterns and make connections among algebraic, tabular, and graphical representations with high accuracy in shorter periods of time compared to manual plotting of graphs (Heid, 1995; Interactive Educational Systems Design, 2003). This is particularly important as students find it very hard to understand relationships and translate among representations, because each representation and translation among them requires a different psychological process (Leinhardt, Zaslavsky, & Stein, 1990). For example, in most problems involving graphing, translations from equation to graphs and from graphs to equation are required. Concerning these two translations, movement from graphs to their equation would be a more difficult task because it involves pattern detection. If the latter is in the form of a matching-type item, students check for the equations and do not attempt to move graphs to equations. On the other hand, the graphs in textbooks usually have equal scales, and critical points (e.g., x-y intercepts, axis of symmetry) are usually integers. However, in realistic or real-life problems, students are faced with unequal scales and rational or decimal critical points, and they find it difficult to interpret results on the graphs. Partial views, axes without labels, and individual points on the graphs are other difficulties for students to interpret. However, with utilization of graphing technologies, students are forced to think about such issues, since the equation may be given with or
without the scale, or even with a partial view (Cavanagh & Mitchelmore, 2003; Leinhardt, Zaslavsky, & Stein, 1990).

**Methodology**

**Participants**

Like other national standardized exams in Turkey, students are prepared for university entrance exams mostly in *dershanes*, which is the Turkish counterpart of the cram schools found all over the world. While the high school graduates preparing for the university entrance exam typically attend *dershanes* on weekdays, those attending regular schools attend *dershanes* on weekends and in some instances after school hours. In *dershanes*, students are mostly taught towards standardized tests for solving multiple-choice items. Based on their intended field of study at university, they are required to solve tests including mathematics (mostly arithmetic, algebra, and geometry), Turkish language, history, geography, and philosophy.

The participants in this study were 65 high school graduates preparing for the university entrance examination. The experimental group (EG) consisted of 31 students while the control group (CG) consisted of 34 students. Although there were 35 students in each group at the beginning of the study, the number of participants for each group is defined as the number of students who completed the treatments and all the tests administered. The students’ ages in both groups ranged from 18 to 20. Students in both groups were attending two separate *dershanes*. Thus, students in the experimental and control groups did not know or have a chance to interact with each other. Moreover, in both groups, students’ social and economic statuses were nearly the same. Except for two students in the control group, students in both groups had only one parent working and family incomes were nearly the same. Also, none of the students had a computer at home. The students in both groups were chosen based on the scores that they received from the university entrance examination the previous year and their high school grades. In terms of these, the students that participated in the study were average students.

**Instruments**

*Graphing Achievement Test*

The Graphing Achievement Test (GAT) was developed by the researchers to measure student achievement in graphing quadratic functions. It consisted of 15 items, some of which included sub-tasks: 10 multiple-choice, 3 matching-type, and 2 essay-type. Explanations are required for answers in all items, including the multiple-choice ones. The items in the GAT consist of three task types to assess how students interpret and construct graphs of quadratic functions: translation (translating among algebraic, tabular, and graphical representations of quadratic functions; translating a graph of a quadratic function, parabola), classification (classifying given parabolas using properties and coefficients), and drawing (drawing the graph of a quadratic function, a parabola, provided in algebraic form) (see Leinhardt, Zaslavsky, & Stein, 1990). Furthermore, all items involved the concept of scaling.

The test was checked by three mathematics teachers for level of difficulty, accuracy of wording, and suitability to measure the related objectives in the curriculum. The pre-test consists of the same number of questions with the same context as the other two tests. The only difference between the pre-test and the others is that different numbers are given in every question assessing the same objective. The post-test was piloted with 75 students chosen from three different *dershanes* in Ankara. The students were given 60 minutes to complete the test. Students were interviewed to reveal if the tasks were easily understandable and if there were any incorrect or incomplete questions, and also to assess the reliability and consistency. After this pilot, the test remained unchanged. Cronbach’s alpha was .82.

The GAT was administered as a pre-test, post-test, and delayed post-test (two months after the treatment was over) to both groups. Fifty minutes were given to complete the test. All tasks in the test were scored as 0, 1, or 2 for each incorrect answer, correct answer, and correct answer with complete explanation, respectively. The maximum possible score in all three tests was 58.
Attitudes towards Mathematics Inventory

The Attitudes towards Mathematics Scale (ATMI) (Tapia, 1996; Tapia & Marsh, 2002, 2004) was administered to the experimental group before and after the treatment to assess the students’ attitudes and attitude changes toward mathematics. The ATMI consisted of 40 five-point Likert type items with 28 positive and 12 negative statements: from “strongly disagree” (0) to “strongly agree” (4). Higher scores indicate more positive attitude towards mathematics. Students were given 20 minutes to complete the ATMI. Cronbach’s alpha reliability coefficient of internal consistency was .75.

Attitudes towards Technology Scale

The Attitudes towards Technology Scale (ATTS) was administered to the experimental group before and after the treatment to assess students’ attitudes toward technology use in mathematics lessons. The ATTS consisted of 20 items scored on a Likert-type five-point scale ranging “never” (0) to “always” (4). Higher scores indicate a positive attitude towards technology. The students were given 10 minutes to complete it.

The first 10 items that were taken from the Mathematics and Technology Attitudes Scale (MTAS) (Pierce, Stacey, & Barkatsas, 2007) were about attitude towards computer technology. Cronbach’s alphas for the confidence in using technology and attitude to the use of computers to learn mathematics sub-scales were .79 and .89 respectively. The remaining ten items were taken from the Computer Attitude Questionnaire developed by Knezek and Christensen (1996) for middle and high school students. The scale measures IWB enjoyment and motivation, importance of IWB, and use of IWB to learn mathematics. The corresponding Cronbach’s alpha values for these sub-scales were .80, .84, and .81, respectively.

Procedures

The study was designed as a pre-test, post-test, delayed post-test quasi-experimental control group study in which two different teaching and learning environments were utilized: one with traditional instruction and the other with the incorporation of an IWB and computer sessions. Students in the experimental group were taught in a classroom with an IWB and in a computer laboratory with 16 computers. Connected to a large screen, one of the computers was for teacher use. On the other hand, students in the control group were taught in a classroom with a blackboard and did not have computer access. Other than the IWB and computers, class size and classroom items (seats, desks, etc.) in both cases were similar.

Both experimental and control group students were taught by the second author, who followed the same sequence and content of lessons and solved the very same questions in both classes. The lessons and assignments were in line with the standards and objectives dictated by the MoNE. The content was also aligned with the official textbook of the MoNE for the tenth grade, where the topics are first introduced. The lessons included tabular, graphical, and algebraic representation of quadratic equations/functions, their translations, and their properties depending on the coefficients and discriminants of the equations and the symmetry axis.

After preparing assignments and lessons, two other mathematics teachers’ opinions were sought: One was 35 years old with 10 years of teaching experience and the other was 37 year-old with 12 years of teaching experience. The assignments’ levels are determined by the students’ success from their high school mathematics scores and scores from their exams in the dershane.

To assess the appropriateness of the lesson plans, activities, use of IWB, and computers, the study was piloted with 7 students. All students were asked to write an essay about their feelings and thoughts regarding the learning environment and the lessons. We also carried out an hour-long interview with 2 of the students. While no problems were encountered in terms of IWB and computer usage, some changes (e.g., removing or clarifying the content, increasing the number of solved questions regarding translating among different representation types) were made in the lesson contents.
Data were collected over a period of two weeks. Participants met twice a week and took four hours of lessons, each of which lasted 60 minutes for both groups. In the experimental group, class sessions were followed by computer laboratory sessions, each lasting about 60 minutes. Furthermore, interviews were carried with 6 students chosen based on their scores in pre- and post-tests to learn more about (i) whether the IWB created a difference in their learning, (ii) if and how the IWB changed the students’ opinions about mathematics and technology, and (iii) what would be the advantages and disadvantages of using the IWB. Three of these students (i.e., Students A, B, and C) got zero points from the pre-test of the GAT. Their scores for attitude towards mathematics and technology were also lower prior to the study. They expressed that they did not know or have any idea about quadratic equations and functions. The other three (i.e., Students D, E, and F) were average students according to their pre-test scores. They had an average or above average attitudes prior to the study. Despite their differences, all of these 6 students increased their performance in the GAT and attitudes significantly after the treatment. The interviews were video-recorded.

**Treatments for the experimental and control groups**

For the experimental group, lecture notes were given to the students in the lessons and also reflected on the screen of IWB. However, the control group students needed to take notes in the lessons and so sometimes the lectures took longer as compared to the experimental group. Students worked with NuCalc graphing software (Avitzur et al., 1998) in the computer lab. In order to draw a graph with NuCalc, the user only needs to enter the equation and press the enter button without the need for any command syntax. The post-lesson computer laboratory sessions for the experimental group helped students to explore and interpret relationships between the knowledge that they learned in the classroom and gave them the opportunity to draw multiple graphs in order to observe and interpret relationships as coefficients in the functions changes. In the NuCalc environment, it is possible to draw multiple graphs with different colors on the same screen so that users can compare graphs by their coefficients. Moreover, zoom in and zoom out options are available in the program and the most powerful feature of the software is its capability to turn a graph of any function into an animation. The software recognizes the letter “n” in an equation as a parameter and graphs can be animated on any variable and on any defined domain. The experimental group was taught about NuCalc before the treatment by the second researcher and it took 60 minutes. During this time students learned coordinate systems, points on the coordinate systems, to plot graphs of linear functions, to draw more than one graph in one coordinate system, to draw graphs of quadratic functions, to draw multiple graphs on the same screen, to graph polynomial functions with any degree, and other functions like sine, cosine, and absolute value.

In the experimental group, the first part of the treatment took place in the classroom where the IWB was present. Before the lessons, the lecture notes were prepared by the teacher (i.e., the second researcher) and were distributed to the students. The lecture notes consisted of the definitions, main titles and subtitles, and examples to be solved together in the classroom. The notes were reflected on the IWB screen. As they did not need to write down the information and definitions, students had more time to listen, discuss, and solve more examples. The graphs and tables were ready to draw and to fill in on the IWB. Such graphs and tables were also present in the students’ notes. Thus, the teacher and the students together constructed the table with values and drew the graphs on the coordinate system. After drawing a couple of graphs, multiple graphs with different colors were drawn on the same coordinate system. The students had a chance to compare and contrast the graphs and interpret their properties, which are connected to the solution sets. They also learned to draw graphs of the following types of functions:

\[
y = x^2 + a, \quad y = (x + a)^2, \quad y = (x + a)^2 + b, \quad \text{and} \quad y = ax^2 + bx + c \quad (a,b,c \neq 0) \quad \text{starting with} \quad y = x^2.
\]

After the lessons with the IWB, the computer laboratory session took place. Each student had one computer to practice on and the researcher had the main computer, which was connected to a large LCD TV screen. All the computers were pointed towards the screen so that students had a chance to check their answers from the main computer screen reflected on the TV screen. With the activities in the laboratory, students continued doing algebraic, tabular, and graphical representations and translation problems using NuCalc. However, this time the researcher showed only the first activity and the whole class discussed the results together; then the students did the rest of the activities by themselves, individually. Moreover, before drawing the graphs, students were expected to draw by hand first and then to check from the screen. With these activities, students were expected to practice the properties and use the computer as a checking tool.
In the control group, on the other hand, students studied the topic through traditional direct instruction. That is, students took notes from the blackboard and solved the question asked. The teacher explained the concept by writing down the definitions and properties and waited for the students to write. This writing part took more time and so the control group solved fewer problems compared to the experimental group. Nevertheless, they mainly solved the same problems in the same order. More time was spent on drawing and waiting for students to draw the graphs and less time was spent on discussion. Moreover, the control group solved slightly fewer examples compared to the experimental group because the writing and drawing took longer.

**Results**

**Student achievement in graphing and interpreting quadratic functions**

To compare changes in students’ GAT from pre-test to post-test and to delayed post-test between experimental and control groups, pre-test, post-test, and delayed post-test scores for both groups were calculated (see Table 1). Mean test scores indicate that students in both the experimental and control groups have made significant content knowledge gains from pre-test to post-test, and lost a portion of it between post-test and delayed post-test.

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics regarding experimental and control groups’ pre-test, post-test, and delayed post-test scores for the GAT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Pre-test</strong></td>
</tr>
<tr>
<td><strong>M</strong></td>
</tr>
<tr>
<td>EG (n = 31)</td>
</tr>
<tr>
<td>CG (n = 34)</td>
</tr>
</tbody>
</table>

Conducting mixed ANOVA, Mauchly’s test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 20.5, p < .001$); therefore, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .81$). The results show that students’ performances significantly changed across times of testing (i.e., pre-test, post-test, and delayed post-test), $F(1.62, 101.97) = 583.97, p < .001, \eta^2 = .90$. Furthermore, contrasts revealed that the students performed significantly better in the post-test relative to both the pre-test, $F(1, 63) = 783.35, p < .001, \eta^2 = .92$, and the delayed post-test, $F(1, 63) = 12.5, p = .001, \eta^2 = .17$.

There was a significant interaction effect between time of testing and groups, $F(1.62, 101.97) = 16.9, p < .001, \eta^2 = .2$. This indicates that student performance across time of testing was dependent upon the presence or absence of the treatment. Contrasts comparing students’ scores in the EG and CG across the three times of testing revealed significant interactions when comparing EG and CG students’ scores in the delayed post-test compared to the pre-test, $F(1, 63) = 24.06, p < .001, \eta^2 = .28$, but not in the post-test compared to the delayed post-test, $F(1, 63) = 0.04, p = .85, \eta^2 = .001$. This suggests that although in both the EG and CG students’ performance increased from pre-test to post-test and then decreased from post-test to delayed post-test, the rate of decrease was about the same and the rate of increase was different, with students showing a greater rate of increase in the EG. The results indicate that only the groups (i.e., intervention factor) exerted a significant effect on students performances, $F(1, 63) = 22.54, p < .001$. In other words, the EG achieved better than the CG, averaged across the pre-test, post-test, and delayed post-test.

**Students’ attitudes toward technology and mathematics**

Table 2 shows mean scores for the attitude scales for mathematics and technology (i.e., ATMI and ATTS) administered only to the experimental group before and after the treatment. Results show that the students developed a more positive attitude toward technology and mathematics after the treatment. For example, although 59.4% of the students were undecided about the statement “I concentrate better in class when a whiteboard is used to deliver instruction” prior to the study, 93.8% of the students either agreed or strongly agreed with it after the study.
Moreover, although most of the students were undecided about “Interactive whiteboard and computer would help me to learn mathematics better” prior to the study, 80% of the students either agreed or strongly agreed with it at the end.

Table 2. Descriptive statistics regarding the ATMI and ATTS scores for the experimental group (n = 31)

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th></th>
<th>Post-test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>ATMI</td>
<td>33.37</td>
<td>7.90</td>
<td>65.78</td>
<td>3.58</td>
</tr>
<tr>
<td>ATTS</td>
<td>55.09</td>
<td>19.27</td>
<td>103.18</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Regarding the ATMI, on the other hand, students’ positive attitudes and feelings could be concluded. To illustrate, prior to the study, for example, regarding Item 21, “I feel comfortable in mathematics lessons,” no students selected strongly agree, 6.3% selected agree, 31.3% selected undecided, 37.5% selected disagree, and 25% selected strongly disagree, showing that most of the students felt uncomfortable in their mathematics lessons. On the other hand, after the treatment, strongly agree increased to 37.5%, agree increased to 50%, and undecided decreased to 12.5%, showing that the students were more comfortable in mathematics lessons when IWB and computer technologies were used.

The related sample t-test was run in order to examine whether students in the EG developed more positive feelings and attitudes toward mathematics and technology. The results revealed that the EG post-ATTS scores were significantly higher than those of pre-ATTS: t(31) = 24.303, p < .001.

Students’ performance in drawing and interpreting graphs of quadratic functions

Although students in both groups improved their achievement in drawing and interpreting graphs of quadratic functions, it was found that students in different groups had different solution strategies. Considering Items 1, 2, and 3, students in the CG were good at drawing graphs from equations by following the steps/procedures given in the lessons (i.e., finding ordered pairs, plotting them on coordinate system, and drawing the equation). However, considering Items 4, 7, 8, 9, and 10, they could hardly transfer/integrate other knowledge like the discriminant, symmetry axis, and coefficients. On the other hand, the students in the EG were also good at drawing graphs considering Items 1, 2, and 3. However, along with following step-by-step instructions, they also used translation, coefficient of functions, discriminant, and symmetry axis for Items 6, 7, 8, 9, and 10. For example, in Items 7 and 8 students were expected to relate a given graph with its equation among the alternatives (see Figure 1).

Table 3. Frequencies of student answers for Items 7 and 8 in pre-test, post-test, and delayed post-test of the GAT in control and experimental groups

<table>
<thead>
<tr>
<th>Item 7</th>
<th></th>
<th>Item 8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CG</td>
<td>EG</td>
<td>CG</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Pre-GAT</td>
<td>20</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Post-GAT</td>
<td>1</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Delayed Post-GAT</td>
<td>8</td>
<td>11</td>
<td>14</td>
</tr>
</tbody>
</table>

Note. 0: incorrect; 1: correct; 2: correct with complete explanations.

Regarding Item 7, while most of the students had difficulty in choosing the correct equation in the pre-test, the majority of them were able to find the corresponding equation. However, in the delayed post-test, the number of students who found the correct answer in the CG decreased more compared to that in the EG (see Table 3). Most of the CG students attempted to solve it by trying to draw the corresponding graphs for each equation in the alternatives. The rest tried to substitute 0 for x in each alternative and compare with the given graph. Two students found the correct equation by substituting −1, 0, and 1 for x and finding the coefficients a, b, and c in ax²+bx+c. On the other hand, 10 students in the CG used an x² graph to find the desired equation. They translated the graph 1 unit downwards and found the correct choice. The rest of the students preferred using the y-intercept to find the coefficient c. They reasoned that since the graph passes x = −1, then c = −1, and thus they eliminated alternatives A and D. They then substituted 1 and −1 in the remaining alternatives and found the correct answer. On the other hand,
while Item 8 assesses the same knowledge as Item 7, it was more difficult since the graph did not cross the x-axis and students could not easily eliminate the alternatives. Thus, most of the students had difficulty in finding the correct answer (see Table 3). The CG students mostly choose A, a distractor. They easily eliminated alternatives B, C, and E, where the constants are −1 and the graphs passes through 1 on the y-axis. After eliminating these alternatives, they could choose between A and D. Most of them chose A, since it was more familiar to the students and they did not take the effect of the leading coefficient in the equation into account. On the other hand, students in the EG also eliminated alternatives B, C, and E for the same reasons. However, most of them compared the given graph with the one in Item 7 and concluded that the leading coefficient needed to be larger than zero as the graph was closer to the y-axis. Also, 7 students picked certain points on the graph and tried substituting those in the alternatives to find the correct one. Two EG students’ answers for Items 7 and 8 are provided in Figure 1. Those students approached the problem more visually rather than from an algebraic manipulation point of view as they used translation of the graph of y = x^2 to relate the given graph with an equation.

In the post-tests, the EG did not depend on the step-by-step solution; they used translation among graphic, tabular, and algebraic forms. They also considered the effect of coefficients on graphs and properties of quadratic equations like the discriminant and symmetry axis. They provided explanations about the graphs that they drew. This provides evidence that the EG students used critical properties related to quadratic equations/functions and their understanding was deeper.

EG students did better in interpreting items involving classification and translation tasks as compared to those in the CG. As emphasized by Leinhardt, Zaslavsky, and Stein (1990), the major difficulty for students of all ages is the translation between algebraic, graphical, and tabular representations. Students find it very hard to understand relationships and translate among representations because each representation and translation among them requires a different psychological process. For example, most items in the GAT required translations from an equation to graphs and from graphs to an equation. Concerning these two translations, movement from graphs to their equation would be a more difficult task because it involves pattern detection.

For drawing graphs from the algebraic form, students followed the steps (i.e., finding ordered pairs, plotting them onto coordinate system, and drawing the equation). For finding the equation from a given parabola, students had more difficulty. For matching-type items, students attempted to draw graphs for the equations without attempting to make a connection between the graphs and the equations. This conclusion had some similarities to Barton’s (1997) study, where he also found that the group that used technology made more connections to their previous knowledge.
(i.e., transfer), suggested better (or well-formed) explanations, and made better predictions (or conjectures) as compared to the CG.

**Students’ views about graphing**

All six students interviewed expressed that they developed positive feelings and attitudes toward graphs and graphing in mathematics after IWB use and expressed that their explanation and reasoning skills improved. They indicated that they could comment on and make interpretations regarding a given problem about the quadratic functions.

Before taking this lesson, I wouldn’t even guess what the shape of a quadratic function looked like. But after the lessons I can now visualize the drawings [i.e., the graphs] even when we talk. I can explain now that the coefficient of \( x^2 \) indicates whether the parabola is closer to or far away from the \( y \)-axis and the constant shows the \( y \)-intercept and each unit that we add the \( x \), the parabola changes the place on the \( x \)-axis. Moreover, I learned the relation between the discriminant and the \( x \)-intercepts of the parabola. So in any question just by knowing this I believe that I can draw any graph without calculation and in the test I can find the correct answer. So I am not afraid of seeing a graphic in the test now. (Student A) [italics added]

At first, I was just thinking that graphing consists of a given algebraic expression and drawing it and vice versa. However after the lesson, I realized that I was spending too much time drawing the graph and if I could not find the intercepts I was giving up drawing. Also, if the [multiple-choice] question required writing the equation of a parabola I was checking the equations given in the alternatives and drawing the matching graph for each to see if any one of those matches with the graph given in the question. However, I learned that drawing [graphs] includes more than this. Now I can use different representations: tabular [i.e., numerical], graphical, and algebraic to find the correct answer. I can easily translate among different modes of representations. This takes less time and sometimes I can find the correct choice without writing or calculating anything. (Student F) [italics added]

One of the students (Student D), on the other hand, explained her views by using her hand as the shape of a parabola and moving her hand as if the coefficients changed. She was very enthusiastic about describing what each shape means based on the changes in the parameters. While talking about reasoning skills, Students D and E explained that they could also apply their knowledge gained during the study to various graphs other than quadratic functions. They indicated that they could solve problems involving graphs of the first and third degree as well as some

**Students’ views about learning with IWB**

Analysis of student interviews revealed that students found doing mathematics with IWBs more interesting \((n = 4)\) and fun \((n = 5)\). Students mostly mentioned the visualization \((n = 6)\) and the efficiency and ability to save time for discussion and examples \((n = 5)\). Except for Student B, the others had positive comments about the IWB.

The IWB is good only for saving time. Since I do not have to write down all the things on the board I have more time to solve more examples. Other than that I do not think that the IWB helped me to understand better. Because I cannot easily understand things in class; in order to understand them I need to study by myself. I need to apply what I learn. (Student B)

He explained that the computer and the NuCalc software helped a lot more with drawing and translation. He could study alone with the computer so that he could think more about examples and connect them with the lesson. In this sense, he thought he would benefit more from the computer (i.e., the software) than the IWB, yet he found that the IWB made the lessons more interesting. Unlike Student B, on the other hand, Student C insisted that the IWB should be used in every mathematics lesson, basically because of the fact that students can use the extra time to focus on the solutions instead of trying to write down everything on the board.

I cannot imagine a lesson without the IWB. It is great because the teacher’s notes are reflected on the board and we have copies and so we have time for more examples, for more talking. We especially discussed why and how questions helped me to improve my learning. In other lessons, we try to write down everything on the board, and if we are late in finishing writing we skip the example without really understanding it. Also, I can say about the IWB that I especially love using different colors in drawings of graphs. Seeing different
graphs with different colors in one coordinate system helped me to understand the relation between the changing coefficients and the corresponding graphs. Now I have started to like mathematics. Before using the IWB, I sat at the back of the classroom and did not want the teacher to ask me any questions. I just listened to the teacher without thinking. But now, the IWB makes me think about the lesson, because I start to have fun and that makes me interested in the lesson. And I found myself waiting for another question, checking my answers. I think the IWB motivates me in the lesson. (Student C)

**Students’ views about learning with computers**

Analysis of interview data indicated that students found that drawing graphs is quicker \((n = 4)\) and more fun \((n = 3)\) with NuCalc. They also mentioned the visualization \((n = 3)\) and immediate feedback \((n = 3)\) that the software enabled. Prior to the study, some students indicated that they had difficulties using computers. However, all students interviewed expressed that they found it very easy to use the NuCalc software. As Student-A expressed, “Anyone who can type the equation and press the enter key can use the program; there is no need to know anything about the computer itself.” During the computer sessions, however, some students had difficulty entering the correct formula, particularly those including brackets. However, they either asked for help or tried to enter the equations again by themselves. All of the students expressed that they loved being able to draw a graph in a flash.

Drawing a graph with one click is amazing. After a couple of questions before I hit the enter key, I tried to visualize the shape of the equation: where it passes, downward or upward, intercept points, and after a couple of tries a lot of the time I guessed the correct shape. This was wonderful for me because at first I did not know what a parabola looked like but now I can draw them in my mind. This will be very helpful for me while selecting the correct equation and correct graph. (Student B)

Concerning attitudes and feelings, all students expressed that they were very happy to learn the subject with computers.

I am very enthusiastic when sitting in front of a computer, checking my own answers, interpreting my graphs and results. I like to do the lesson by myself, because I learn more and also learn more easily. It arouses my curiosity. I try to guess the next step and I see all representations on one page. Comparing the IWB and the computer, the computer increases my learning a lot more. I can think more clearly when I use a computer to do mathematics and I want to do more graphs using the computer. (Student B)

When I first sat in front of the computer, it scared me a lot, because I did not know anything about computers. It looked so complicated until the teacher explained the first example. Before doing the activities, we practiced. After this work I feel comfortable using the computer. I enjoyed drawing graphs at the computer. It made mathematics more interesting and fun. (Student C)

**Discussion and conclusions**

The results indicated that both EG and CG students’ performances increased from pre-test to post-test and then decreased from post-test to the retention test. The rate of decrease was about the same; however, the rate of increase was different, with students showing a greater rate of increase in the EG, where students were instructed in an IWB- and computer-supported environment. Furthermore, the treatment positively affected students’ attitudes towards technology and mathematics in EG. The students in the EG were enthusiastic about the technology use and were actively involved in the lessons, asking questions and making connections between what they learned before. Students in the CG got bored easily after drawing only a couple of graphs as they thought all such graphs were nearly the same. Moreover, reasoning and interpretation skills of CG students did not improve as much as those in the EG.

It is more likely that students are willing to engage with learning tasks and motivate themselves to learn materials when IWBs are employed (Painter, Whiting, & Wolters, 2005; Smith, Hardman, & Higgins, 2006), but nevertheless it would be fair to say that instructional technology is only as good/effective as why and how we use them. Confirming Condie and Munro (2007, p. 5), the results of this study support that better IWB integration in mathematics classrooms should consider supporting IWBs with appropriate software and, in particular, dynamic environments that the teacher might use for demonstration and whole-class discussion purposes and that students might use for investigations or inquiry. Furthermore, such tools are only good in the hands of a teacher who actually
knows the responsible use of technology. It is the skills and professional knowledge of the teacher that determines how much value would be gained from the use of IWBs in the classroom (Clarkson, 2011; Higgins, Beauchamp, & Miller, 2007).

Various studies (e.g., Beauchamp, 2004; Glover & Miller, 2001; Hennessy, Fung, & Scanlon, 2001; Şad & Özhan, 2012; Wood & Ashfield, 2008) provide evidence for the relationships among students’ views, perceptions, attitudes and IWBs. Generally speaking, using IWBs or similar technology in the classroom is perceived as a tool that enhances interests, motivation, and enthusiasm, as well as providing prolonged engagement with the learning materials. Consistent with the literature, the present study revealed that students in the EG were enthusiastic about the use of technology. They were actively involved, asking questions and making connections with the knowledge that they had learned before. In the CG, however, students got bored after drawing a few graphs and they thought that all the graphs were nearly the same (i.e., as if they were doing the same things over and over again). Furthermore, students interviewed described the IWB and computers as enjoyable, interesting, and more fun, although they were undecided or did not agree that the IWB and computer could help them to learn mathematics prior to the treatment. The students indicated that they concentrated better in class when the IWB was used and that the IWB and computers helped them to learn mathematics better. Similar results were reported by others (e.g., Beauchamp, 2004; Glover & Miller, 2001; Hennessy, Fung, & Scanlon, 2001; Wood & Ashfield, 2008). Furthermore, data indicated that while no students selected “strongly agree” or “agree” for Item 21 in the attitude scale prior to the study, these answers increased considerably (e.g., 37.5% for “strongly agree”) after the study.

In this study, the use of computer-based graphing allowed learners to practice newly acquired skills or reinforce previous knowledge. For finding an equation from a given parabola, students had more difficulty. For matching-type items, students attempted to draw graphs for the equations without attempting to make a connection between the graphs and the equations. This result is similar to those of Barton (1997), who also found that the group that used technology made more connections to their previous knowledge (i.e., transfer), suggested better (or well-formed) explanations, and made better predictions (or conjectures) as compared to the CG. The reason for this difference might be the visualization and experimentation brought by the technological environment (Archavi & Hadas, 2000). The IWB and NuCalc created an environment in which students could visualize the graphs together in all three forms. Students see the changes in graphs by changing coefficients and this helps them visualize the translation of graphs to desired units (especially the animation and drawing of multiple graphs) and interpret the changes of coefficients on graphs. This conclusion can be drawn from the students’ GAT results and interviews, as well. On the other hand, in the CG the teacher drew graphs on the board and informed students about the properties and how to draw the graphs. Then, similar to the other class, the teacher drew multiple graphs in one coordinate plane to show the translation. However, it can be concluded from the students’ answers from questions that involved translation that this was not enough for the students to develop their reasoning and interpretation skills.

According to Higgins, Beauchamp, and Miller (2007), “The research literature has yet to demonstrate the direction that teachers need to move to ensure that the proven changes the IWB can bring about in classroom discourse and pedagogy are translated into similar and positive changes in learning” (p. 221). In the present investigation students in the experimental group (EG) had access to IWB and NuCalc software by which multiple representations could be used. Students were also situated in a classroom environment where they had more opportunities to discuss and interpret graphical representations compared to those in control group (CG). Thus, teachers should realize the significance of students’ participation in learning activities and they should give primary emphasis to activities that enable students to become actively involved in classroom participation so that students perceive representations as processes rather than end products (Greeno & Hall, 1997). Specifically, if students take part in activities to construct tables and graphs that enable them visualize the changes in their variables, and if they use these representations to justify, explain, and interpret the meaning between the patterns of variables, then they approach representations as thinking tools by which they can build their conceptual understandings. However, if students perceive the purpose of activities as being only to write equations or construct tables and graphs for completing their assignments, they might not develop the view that representations are tools for communicating and understanding sophisticated relationships among concepts. Thus, when the teacher places primary emphasis on teaching technical and conventional representations for succeeding on a test rather than for communicating and understanding concepts, students might not find opportunities to apply what they learn using forms of representations as resources in thinking and communicating in the classroom context.
In conclusion, this study provides further evidence that effective integration of technology into classroom instruction can positively impact students’ motivation, engagement, and interest in learning while fostering an active, explorative, and investigative style of learning resulting in improved knowledge in mathematics. In this context, IWBs can make a difference in students’ achievement, but the extent of this difference seems dependent on how we use them. Last but not least, the authors of this paper believe that teachers and schools should make good use of what can be a significant investment; effective uses of IWBs should be more thoroughly and robustly explored.

Limitations

There are two main limitations to this study: the duration of the study and the two different settings (i.e., lecture with IWB and computer lab with NuCalc) in the EG. Although the delayed post-test provides more evidence regarding the retention of the achievement gains, the improvement in students’ achievements in favor of those in the EG and their strong interest and positive attitudes towards learning with the IWB and graphing technology might be attributable to some novelty factor or Hawthorne effect. Thus, the findings and conclusions of this study should be interpreted accordingly. Additional research efforts with longitudinal studies would give a better picture of how students react and what gains they might have from learning mathematics in similar settings (Digregorio & Sobel-Lojeski, 2010; Hennessy & London, 2013). On the other hand, it is difficult to isolate treatment effects from each other. Thus, when interpreting the results of this study, the learning environment in the experimental group should be conceptualized as a technology-supported (i.e., IWB and NuCalc) student-centered collaborative inquiry.

Acknowledgements

Ayhan Kursat ERBAS is supported by the Turkish Academy of Sciences through the Young Scientist Award Program (A.K.E./TÜBA-GEİP/2012-11).

References


Effects of Verbal Components in 3D Talking-head on Pronunciation Learning among Non-native Speakers

Ahmad Zamzuri Mohamad Ali*, Kogilath Segaran and Tan Wee Hoe
Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjong Malim, 35900, Perak, Malaysia // zamzuri@fskik.upsi.edu.my // m20102001398@siswa.upsi.edu.my // whtan@fskik.upsi.edu.my
*Corresponding author

(Submitted March 14, 2014; Revised June 30, 2014; Accepted July 25, 2014)

ABSTRACT

This study was designed to investigate the benefit of inclusion of various verbal elements in 3D talking-head on pronunciation learning among non-native speakers. In particular, the study examines the effects of three different multimedia presentation strategies in 3D talking-head Mobile-Assisted-Language-Learning (MALL) on the learning achievement of students with low English pronunciation skills. Pre-test and post-test scores were utilized to determine the students’ overall performance. 60 college students with low pronunciation skill were involved, in which they were divided into three equal groups of 20 students. Scores obtained were analyzed statistically with one-way analysis of covariance (ANCOVA), with the pre-test scores as covariate. The findings revealed that 3D talking-head with spoken text and on-screen text MALL has significant contribution in retaining the correct pronunciation acquisition in comparison with 3D talking-head with spoken-text alone MALL and spoken text with on-screen text MALL. Therefore, it may be concluded that multiple sources of verbal information to identify and decode language input are advantageous for effective pronunciation learning, specifically among non-native speakers.

Keywords

Animation, Language-learning, Talking-head, MALL, Non-native

Introduction

Discussion of animated characters typically revolves around cartoons, special effects, and movies. However, it appears to be many positive results indicating that animation has an important role in the field of education as well (Balasubramaniam, 2012; McMenemey & Ferguson, 2009; Doyle, 2001). Recently, animations have been included as a part of multimedia learning aid in numerous subject matters, including in language learning (Lin & Tseng, 2012; Kayaoğlu, Dağ Akbaş & Öztürk, 2011). In the past, students who learn English as a second language depended heavily on printed text and audio materials such as cassette tapes, audio CDs and radio broadcast (Xiao & Jones, 1995). Speech and language technology evolved under the assumption that speech was merely auditory event (Massaro, Liu, Chen & Perfetti, 2006). Nevertheless, numerous research findings revealed that, students’ understanding of the language are also influenced by speaker’s face and accompanying gestures, in addition to the actual sound of the speech (Massaro et al., 2006). Therefore, inclusion of animated character that functions as pedagogical agent in language learning instructional aids seems meaningful.

An appropriate teaching approach is important when learning a second language with the assistance of animated character (Massaro, Bigler, Chen, Perlman & Ouni, 2008). It is very meaningful for overcoming problems faced by students who need to practice clear pronunciations of new words pertaining to a second language, which is different from their first language (Cook, 1996). Notably, numerous instructional approaches were undertaken in establishing English as a second language being acquired globally, which eventually resulted in the introduction of Computer-Assisted-Language-Learning (CALL) and Mobile-Assisted-language-Learning (MALL). In CALL or MALL, the embodied agent, or talking-head animation, becomes the prominent virtual aid for teaching pronunciation, vocabulary, articulation and so forth (Wik & Hjalmarsson, 2009). Generally, talking-head animation acts as a visual character that functions by saying a word or telling a story to the students (Dey, Maddock & Nicolson, 2010). The talking-head character is limited in a setting that the display on the screen only shows the section from the top of its head to the shoulders (Dey et al., 2010). In addition, talking-head animation was developed by combining the principles of linguistics, pedagogy and replete with a good audio system that is capable of helping students to optimize their pronunciation skills (Massaro, 2003).
3D talking-head

Nowadays, there are increasing educational research interests on the animated pedagogical agent in aiding language learning. (Atkinson, 2002; Baylor & Ruy, 2003; Moreno & Mayer, 2000). Pedagogical agents are animated characters designed to function in educational settings to facilitate learning (Shaw, Johnson & Ganeshan, 1999). Some regard the agents as talking-heads which feature speech, facial expressions, and gestures to implement pedagogical strategies (Graesser, Chipman & King, 2008). The talking-heads may be featured as virtual tutors or teachers in language learning applications, supporting various facets of the learning process, including read aloud and practice conversation (Busa, 2008).

Acquiring a new language could be challenging matters to people from all walks of life, especially adult learners (Tan, Lin & Wang, 2013). In a study that evaluated synthetic and natural Mandarin visual speech, Chen and Massaro (2011) revealed that 3D talking-head can improve the learning process of new language acquisition. Massaro (2006b) identified potential of animated virtual tutors in language learning, specifically for individuals who possess language learning disabilities and require extra instruction. Nonetheless, the requirements have hardly been fulfilled due to the shortage of professionals and skilled teachers who can offer individual or personalized attention (Massaro, 2006a). Teachers commonly used books and conventional media in their attempts to fulfill the personalization needs, but the outcomes were not promising because those media are not customized according to individual learners (Massaro, 2006a). According Massaro (2006b), the face is an indispensable component of a human body, especially for conveying a correct message. The combined movements of lips, tongue, and jaws deliver visual information that could increase the lucidity of aural understanding, even in a noisy situation (Massaro, 2006b). Therefore, visual speech information can play an essential role in assisting learners to differentiate words, which are difficult to distinguish when referring to aural information alone (Jesse & Massaro, 2010).

The gestures, facial expressions and the associated emotions made by a speaker would reinforce the speech (Massaro, 1998). In an experiment carried out by Liu, Massaro, Chen, Chan and Perfetti (2007) to examine the use of visual speech in Chinese pronunciation through a web-based learning program revealed that visual speech has significant advantages in improving learners’ pronunciation in comparison with audio material. However, issues related to the use of text as verbal support in talking-head application were left unsolved. As justified by Ahmad Zamzuri and Kogilathah (2013), excluding text in pronunciation learning might cause difficulties among learners in identifying the syllable breaks for proper pronunciation, specifically the non-native speakers. Hence, research to identify the ideal solution of verbal support in applications that feature talking-head animation seems important.

On-screen text with spoken text

The theoretical framework of this study was mainly grounded on Mayer’s cognitive theory of multimedia learning. According to Mayer’s cognitive theory of multimedia learning, human processes information through dual channels, namely the visual channel that processes visually represented materials and verbal channel that processes audio materials (Mayer, 2001). Mayer (2001) believes that human understanding occurs when learners are able to mentally integrate visual and verbal representations of a subject matter as both channels are being activated simultaneously. There are three assumptions highlighted in Mayer’s cognitive theory of multimedia learning. As discussed earlier, first assumption states that multimedia learning is dual channel activities, which are visual-pictorial channel and auditory-verbal channel. For instance, in the 3D talking-head language learning condition, animation will be processed in the visual-pictorial channel and the pronounced word will be processed in the auditory-verbal channel.

The second assumption is limited capacity, in which each channel in the human cognitive system has limited capacity in processing information. Therefore, designing pedagogically effective multimedia aids for language learning which grounded on theories, continuously been an important issue (Kim & Gilman, 2008). Taking this into consideration, according to Mayer (2001), students able to acquire knowledge better from animation and narration than from animation, narration and on-screen text. Moreover, when image and text both presented visually, the visual channel will be overloaded, besides the redundancy effects due to dual verbal information (Mayer, 2001). Therefore, if the text is presented in audio form alone, it can be processed by the verbal channel, while the visual channel processes the visual information (Mayer, 2001). In some respects, when on-screen text located together with talking-head, students’ demand on the visual modality could lead to split-attention effects (Kim & Gilman, 2008; Craig,
Gholson & Driscoll, 2002). The reason for this effect is that learners might focus their visual attention on animation, rather than on the on-screen text (Kim & Gilman, 2008; Craig, Gholson & Driscoll, 2002).

However, question arises; do these principles apply for 3D talking-head linguistic learning aid? Whereby, text might be helpful in assisting learners in determining the syllable break for correct pronunciation. The study reveals that visual texts can be more preferable under certain conditions, and one of such conditions is instructional pacing. By examining the effects of instructional pacing and text modality on cognitive load and performance, Stiller, Freitag, Zinnbauer and Freitag (2009) have proven that the learner paced visual text instruction was the most efficient condition. Such condition could be applied to applications that feature 3D talking-head with audio and text. Adding to this, learners would pronounce a word accurately if they know the word stress which can be shown in syllable breaks. The correct number of syllables must be created to clearly pronounce the stress pattern of a word (Jones, 2011). Thus, syllable breaks should be placed as text in learning applications.

Finally, the third assumption is active processing, in which learners are involved in active processing in the channels, which includes media selection (verbal and visual), organizing the media into the verbal and visual mental model and finally integrating them with preexisting knowledge which results in meaningful schema acquisition. This happens when corresponding verbal and visual representations are in the working memory at the same time (Mayer, 2001). The issue of integrating visual and verbal information in order to retain it in the long term memory is important in talking-head language learning condition. The application must have the capability in assisting learners to integrate the visual form of the 3D talking-head with facial expression and lip movement with the spoken text or spoken text complemented with on-screen text. Likewise, they are able to store the knowledge acquired from the sensory memory (listening and watching the 3D talking-head) and working memory (integrating 3D talking-head action and the pronounced word) in the long term memory and apply it precisely when required.

**Facial expression and lip synchronization**

Facial expression has been recognized as one of the key concerns to make the language learning efficient and robust when talking-heads are used (Wik & Hjalmarsson, 2009). Research conducted in neuroscience, cognitive science, and psychology justify that emotions which are formed through facial expressions have a significant role in capturing attention, planning, reasoning, learning, memory, and decision making (Picard, 1997). Emotions also play an important role that influences perception, cognition, and creativity (Johnson, Rickel & Lester, 2000; Picard, 1997). In the instructional design for pronunciation learning, gestures and facial expressions have been identified as important features that would facilitate learning (Brown, 2007). In addition, Sime (2006) has determined that facial expression and gestures are two forms of non-verbal communication that can make the classroom learning interesting and motivate learners to engage in the classroom activities actively. Facial expressions particularly could enhance learning capability, such as the ability to recall information (Allen, 2000; Lazaraton, 2004). This would lead to the retention of knowledge captured by learners when learning pronunciation.

In terms of teaching pronunciation, Rodgers (2001) introduced novel methods and methodological prediction, e.g., full-frontal communicativity which is one of the 10 scenarios that engages all aspects of human communicative capacities. Meanwhile, lip synchronization or lip sync is recognized as one of the key features of talking-heads (Lun, n.d.). As English is a language that depends heavily on lip shape, tongue position, teeth position, jaw movement and air flow (Baxter, 1993), the process of learning pronunciation could be practiced by reading the lip sync (Sumby & Pollack, 1954; Benoît & Le Goff, 1998).

**Method**

**Research objective**

Based on the literature overview above, the main objective of this research is to identify if a 3D talking-head with spoken text and on-screen text has significant contribution in retaining the correct pronunciation acquisition in comparison with 3D talking-head with spoken text alone and spoken text with on-screen text. The hypothesis derived from the discussion is as follows:
Ha1. 3D Talking-head with spoken text and on-screen text MALL has significant contribution in retaining the correct pronunciation acquisition in comparison with 3D talking-head with spoken text alone MALL and spoken text with on-screen text MALL.

Teaching material

The multimedia presentation used in this application was mainly based on human computer interaction design principles. The interface design of the MALL, which is the medium of choice, will be simple and the usage of heavy graphics is less, since considering the download time and the launching of the application. The application will start with a welcoming screen. The content of the welcoming screen would be the title of the application, and a button to continue to the next screen. In the following screen, there will be ten words option (Table 1) for students to choose from for pronunciation practice.

<table>
<thead>
<tr>
<th>Full word</th>
<th>Word with syllable break</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aegis</td>
<td>ae.gis</td>
<td>the protection, backing, or support of a particular person or organization</td>
</tr>
<tr>
<td>Archipelago</td>
<td>ar.chi.pe.la.go</td>
<td>a sea or stretch of water having many islands</td>
</tr>
<tr>
<td>Cache</td>
<td>cache</td>
<td>a collection of items of the same type stored in a hidden or inaccessible place</td>
</tr>
<tr>
<td>Cavalry</td>
<td>ca.val.ry</td>
<td>soldiers who fought on horseback</td>
</tr>
<tr>
<td>Foliage</td>
<td>fo.li.age</td>
<td>plant leaves collectively</td>
</tr>
<tr>
<td>Mischievous</td>
<td>mis.chie.vous</td>
<td>causing or showing a fondness for causing trouble in a playful way</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>pro.nun.ci.a tion</td>
<td>the way in which a word is pronounced</td>
</tr>
<tr>
<td>Ubiquitous</td>
<td>u.bi.qui.tous</td>
<td>present, appearing, or found everywhere</td>
</tr>
<tr>
<td>Suite</td>
<td>suite</td>
<td>a set of rooms designated for one person’s or family’s use or for a particular purpose</td>
</tr>
<tr>
<td>Voluptuous</td>
<td>vo.lup.tu.ous</td>
<td>describes a woman who has a soft, curved, and attractive body</td>
</tr>
</tbody>
</table>

There would be instructions on the top of the practice screen to inform the students to select the options. This will be implemented for all three types of prototype. In each prototype there will be a menu which is the list of the difficult to pronounce and commonly mispronounced words. When the students select the particular word, they will be navigated to the text with syllable break and 3D talking-head screen. When the students select the “play” button on this screen, the 3D talking-head will pronounce the word by following the syllable break. Apart from the play button, there are also other navigational buttons on the screen such as, home, next, previous and exit button. The exit button will be displayed in every screen to allow the students to exit whenever they feel like stop using the application (Alessi & Trollip, 2001). Furthermore, application of the exit button suits well for the instruction with drills methodology, which is the pedagogical approach of the study (Alessi & Trollip, 2001). When the students select the “next” button, they will be directed to the full pronunciation practice screen. In this screen, the students can select the “play” button to repeat the 3D talking-head pronouncing the word again. Following that, in the full pronunciation screen, the students can select the “next” button to see the word’s meaning or select the “home” button to go back to the menu screen, or select the “exit” button to see the credit screen and leave the application. This design is similar to the remaining two prototypes, which are the 3D talking-head with spoken text alone MALL and spoken text with on-screen text MALL. The only difference would be, in the practice screen of 3D talking-head with spoken text, will not have the text and as for the spoken text and on-screen text MALL, the screen will only contains audio and text alone. Figure 1, Figure 2 and Figure 3 depicts the difference between the three prototype developed.

Essentially, the 3D talking-head character design was developed into non-realistic facial proportion yet with proper modeled features of human to show the movement of the lip. The 3D talking-head animation was only limited to lip syncing and basic facial expressions. The automated-lip-sync technic using the special plug-in in the 3D software was utilized to synchronize the lip movement with audio. Meanwhile, the non-realism appearance has its own advantages such as, expressions can be exaggerated using non-realistic features, much freedom in designing relatively the human face resemblance, creative touch which makes them more preferable compared to real human
face, users won’t have high expectations towards its realism performance as they would compare to the real human being face appearance, and cost saving since it does not require special equipment and technology to model and animate (Ruttkay & Noot, 2000).

Figure 1. 3D talking-head with spoken text and on-screen text MALL

Figure 2. 3D talking-head with spoken text alone MALL

Figure 3. spoken text and on-screen text MALL

**Procedure**

Our experimental study investigated the effects of three different multimedia presentation strategies in 3D talking-head MALL on the learning achievement of students with low English pronunciation skills. Independent variable is the multimedia presentation strategies of 3D talking-head MALL, whilst the dependent variable is the post-test performance. Basically, the research design is three groups pre-test post-test approach, which all the three groups are experimental group. In detail, group one was assigned with the first strategy which is 3D talking-head with spoken
text and on-screen text, group two was assigned with the second strategy which is 3D talking-head with spoken text alone and group three was assigned with the third strategy which is spoken text with on-screen text. The study was conducted separately for all the groups in the controlled lab environment. Brief explanations were given about the features of the app to ensure the smoothness of students’ exploration throughout the learning process. Prior to the study, the students were required to undergo the pre-test to determine their pre-existing pronunciation skills of the selected words. Students in each group were given 15 minutes of learning time which was deemed suitable based on the literature and observation done during the development phase. Each student was provided with one 7 inch tablet for the mobile learning purpose. Post-test was conducted immediately upon completion of the learning process. The pre-test and post-test are identical, and involve the same words. The same instructor conducted the study session of all the groups. No verbal guide from the instructor has occurred throughout the learning process.

The research participants were 60 college students, whose age ranged from 18 to 20, enrolled in the Diploma in Multimedia Application program and undergoing similar Essential English Communication course. 60 students with pronunciation difficulty were identified through the oral test done prior to the study, which was conducted by a specifically designated linguistic tutor. These sixty students were then divided into three groups with 20 students each, with balanced proportions of gender in each group. Random stratified sampling technique was used to group them equally.

Test instruments

Pre-tests and post-tests were used on the three groups that undergo the three different multimedia presentation strategies respectively. Pre-test and post-test were oral test that only measure the pronunciation performance of the selected words. The tests required students to read the word list given loudly within 15 seconds. Since it consumed time to assess all the three groups immediately, which contains sixty students in total, the test was videotaped and later given to the examiners to assess. By videotaping, also ensures the examiners have ample time to evaluate the students’ performance precisely, whereby, they could replay the video accordingly. Three English lecturers who are expert in linguistic were appointed to assess the students’ performance to ensure the reliability of the result. The examiners assessed the students’ performance based on the oral test score sheet and schema which was adapted from the existing National Education Certificate oral test score sheet. The score sheet and schema was also validated by two English linguistic lecturers. The average grade from the three examiners was used as the final score of the students.

Results

A one-way ANCOVA was conducted in order to examine whether 3D talking-head with spoken text and on-screen text MALL has significant contribution in retaining the correct pronunciation acquisition in comparison with 3D talking-head with spoken text alone MALL and spoken text with on-screen text MALL. The independent variable was the 3D talking-head with three different presentation strategies and the dependent variable was the post-test scores administered following completion of the learning process. Scores of the pre-test administered prior to the commencement of the learning process were used as the covariate.

Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances and homogeneity of regression slopes. From one-way ANCOVA test, Levene’s test for homogeneity of variances was not significant \((p > 0.05)\), and therefore the data do not violate the assumption of equality of error variances. After adjusting for pre-test scores, there was significant difference on post-test achievement of 3D talking-head with spoken text and on-screen text MALL in comparison with 3D talking-head with spoken text alone MALL and spoken text with on-screen text MALL application. \(F(2,56) = 5.65, p < 0.05\), partial eta squared = 0.17, with large effect size according Cohen’s 1998 guidelines (Pallant, 2007) as shown in Table 2. The adjusted mean scores indicate that students in the 3D talking-head with spoken text and on-screen text MALL application obtained a better mean score \((M = 40.70, SE = 0.96)\) than students in the 3D talking-head with spoken text alone MALL application \((M = 36.73, SE = 0.97)\) and spoken text with on-screen MALL application \((M = 36.77, SE = 0.96)\) as shown in Table 3. This clearly suggests that 3D talking-head with spoken text and on-screen text MALL application has significant contribution in retaining the correct pronunciation acquisition in comparison
with 3D talking-head with spoken text alone MALL and spoken text with on-screen text MALL. In summary, the hypothesis derived from the literature overview done is accepted.

<table>
<thead>
<tr>
<th>Table 2. Summary of Tests between-subjects effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Pre</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Error</td>
</tr>
</tbody>
</table>

*Note. R squared .28 (Adjusted R squared = .24).*

<table>
<thead>
<tr>
<th>Table 3. Summary of Descriptive Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
</tr>
<tr>
<td>------------------------------------------</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>3D Talking-head with audio and text MALL</td>
</tr>
<tr>
<td>3D talking-head with audio alone MALL</td>
</tr>
<tr>
<td>Audio and text alone MALL</td>
</tr>
</tbody>
</table>

*Note. Covariates appearing in the model are evaluated at the following values: Pre = 22.91.*

**Discussions and conclusions**

Animation plays potential role in improving human learning process, particularly in promoting profound understanding of the subject matter (Mayer & Moreno, 2002). In line with this, animated pedagogical agent such as 3D talking-head or virtual language tutor, which simulates real human like tutor in the computer, has been created for aiding language learning. The use of 3D talking-head also subsequently improved the learning process of a new language or second language (Chen & Massaro, 2011). Even so, there are dozens of audio and text based talking dictionary available online, the inclusion of animated pedagogical agent seems beneficial for pronunciation learning; specifically for non-native in learning difficult to pronounce English words. English is a language which depends upon; airflow, lip shape, tongue position, teeth position and jaw movement (Baxter, 1993), where the process can be practiced through watching the lip syncing activities (Sumby & Pollack, 1954; Benoit & Le Goff, 1998). Besides that, facial expression is also among the concern to make the language learning more efficient and robust (Wik & Hjalmarsson, 2009). These features, which could be incorporated in the 3D talking-head, might contribute to effective pronunciation learning, specifically among non-native speakers. This research finding further affirms the argument. Whereby, the finding showed that students in the 3D talking-head with spoken text and on-screen text MALL learning condition, outperformed students in the 3D talking-head with spoken text alone MALL, and spoken text with on-screen text MALL, which appears to be the traditional language learning setting.

The belief that speech and language science and technology evolved under the assumption that speech was a solely auditory event seems is not totally true (Massaro, Liu, Chen & Perfetti, 2006). Multiple sources of information to identify and interpret the language input, undoubtedly, promote better pronunciation learning. It could be concluded that 3D talking-head with spoken text and on-screen text MALL, which combines visual information in the form of 3D talking-head and verbal information in the form of spoken text audio and on-screen text display promote better pronunciation learning, instead of one of the element is removed (i.e., talking head or on-screen text). The value of the 3D talking-head is mostly on depicting the lip syncing activities. Whereby, visual information from the movements of the lips enhances lucidity of the audio understanding, specifically in a noisy environment (Massaro, 2006b). Nonetheless, the 3D talking-head character in this study only displays the lip sync in comparison to Massaro’s series of studies that also display the tongue and palate; which were shown by making the skin transparent. Further study comparing these two strategies seems interesting.

Beside the lip sync, facial expression of the 3D talking-head character may also possibly influence the outcome of the study. Emotions which are formed through facial expression have a significant role in capturing attention and facilitate learning (Picard, 1997), which might contribute in enhancing recall ability that would lead to better retention of the knowledge captured (Allen, 2000, Lazaraton, 2004). From the observation throughout the study,
students in the 3D talking-head with spoken text and on-screen text MALL, seemed more excited in exploring the application in comparison with spoken text with on-screen text MALL. This might have been due to the 3D talking-head as pedagogical agents have potential in increasing the motivation and confidence level of the students in learning, since it acts as a virtual tutor. Therefore, further study in looking on the motivation and confidence level aspects and its effects on pronunciation learning should answer this belief.

Besides observing the talking-head in visual terms, the importance of listening in language learning and its role in helping the learner to overcome the barrier of learning a new language has long been acknowledged. Likewise, in this study, spoken text plays an important role in teaching the students to pronounce. When students are exposed to visual information and verbal together, they are able to integrate them with the pre-existing knowledge which results in meaningful schema acquisition in comparison if the information is in verbal condition alone (i.e., spoken text and on-screen text). This happens when corresponding visual and verbal representations are in the working memory at the same time (Mayer, 2001). However, in learning pronunciation apart from 3D talking-head as visual information, visual text or on-screen text also plays an important role. To pronounce the stress pattern of a word clearly, the right number of syllables needs to be produced (Jones, 2011). Therefore, in order to show this, the syllable break was placed as text in the application. From the observation and the result obtained, it was proven that the 3D talking-head with on-screen text has shown significant difference in the interest and score achieved by the students.

On the other hand, the possible redundancy and split attention effects does not manifest any noticeable consequences on the learning performance of students in the 3D talking-head with spoken text and on-screen text condition. This might due to the repetition process which facilitates the learners in developing a more accurate mental model in the cognitive structure for meaningful schema acquisition; as also pointed by Stiller et al. (2009). Even so, the repetition occurs in all the three learning conditions, students in the 3D talking-head with spoken text and on-screen text condition seems benefited more. Multiple sources of verbal information that complemented the visual information might trigger more sufficient active processing in the memory structure for effective pronunciation learning. Possibly, similar learning outcome could be achieved if the learning time was added for the remaining two conditions. However, increased time does not necessarily support learning. Increased learning time may be attributed to other consequences which might impede learning, such as getting bored, decreasing the learning interest, demotivated, etc.

Even though the study revealed that the 3D talking-head with spoken text and on-screen text MALL plays an important role in helping learners distinguish how to pronounce difficult words, the question arises whether the same retention rate can be achieved when examining words in context comparing words in isolation? In this study, students only practice to pronounce the words in isolation. Thus, this study suggests a potential future study should be in comparing words in context besides words in isolation in retaining the correct pronunciation acquisition. Surprisingly, students in the 3D talking-head with spoken text alone MALL, and spoken text with on-screen text MALL learning conditions obtained almost similar mean score. As has been discussed, both of these applications are actually lacking in including one of the elements, which would be crucial for effective pronunciation learning.

Acknowledgements

The authors wish to acknowledge the support of the Ministry of Higher Education and Universiti Pendidikan Sultan Idris, who awarded a RAGS research grant for this study.

References


Applying The CHAID Algorithm to Analyze How Achievement is Influenced by University Students’ Demographics, Study Habits, and Technology Familiarity

Bahar Baran* and Eylem Kılıç

Department of Computer Education and Instructional Technologies, Buca Faculty of Education, Dokuz Eylül University, Uğur Mumcu Street, 35150 Buca, İzmir, Turkey // Department of Computer Education and Instructional Technologies, Faculty of Education, Yuzuncu Yıl University, Zeve Campus, 65080, Van, Turkey // baharbaran35@gmail.com // keylem@gmail.com

*Corresponding author

(Submitted November 20, 2013, Revised September 1, 2014, Accepted October 8, 2014)

ABSTRACT

The purpose of this study is to analyze three separate constructs (demographics, study habits, and technology familiarity) that can be used to identify university students’ characteristics and the relationship between each of these constructs with student achievement. A survey method was used for the current study, and the participants included 2,949 university students from 11 faculties of a public university in Turkey. A survey was used to collect data, and the data were analyzed using the chi-squared automatic interaction detection (CHAID) algorithm. The results of the study revealed that female students are significantly more successful than male students. In addition, the more introverted students, whether male or female, have higher grade point averages (GPAs) than those students who are more extroverted. Furthermore, male students who use the Internet more than 22 hours per week and use the Internet for up to six different aims have the lowest GPAs among all students, while female students who use the Internet for up to 21 hours per week have the highest GPAs among all students. The implications of these findings are also discussed herein.

Keywords

Higher education, Achievement, Gender, Technology, Motivation, Decision tree

Introduction

Because developments in the field of information and communication technology (ICT) have led to a technology-based culture, educators are now educating a new generation of students. Accordingly, one of the important issues to be addressed when designing technology-oriented learning environments is learners’ characteristics. With regard to this issue, the education profession faces two challenges. The first issue is that today’s learners, who have grown up as digital natives and thus are being defined as the Net generation, are vastly different from the generations of learners that preceded current 21st-century learners (Oblinger & Oblinger, 2005; Prensky, 2001). The second issue is that children are experienced in a new form of play, that is, multiplayer computer and video games, and these experiences have shaped their preferences and capabilities regarding learning (Prensky, 2001). Therefore, to provide today’s learners with a better education and a better learning context, educators should consider the characteristics of these new learners.

Gender, working habits, motivation, media use, and book reading, as shown in Table 1, may influence Net generation students’ academic performance. Motivation has been found to be a significant predictor of achievement among students (Ayub 2010; Chan, Wong, & Lo, 2012). In addition, gender has been correlated with achievement (Sarier, 2010; Veenstra & Kuyper, 2004). However, the use of technology by students has controversial results as well (Turner & Corucher, 2013; Junco, 2012). The research studies conducted on study habits did not always reveal consistent results in that while some found significant correlations between study habits and achievement (Yu, 2011), others did not (Olatoye & Ogunkola, 2008). The majority of the studies in Table 1, which have small sample sizes, focused on the dualities between motivation and achievement, and the use of technology and achievement. However, the current study deals with gender, university level and achievement, study habits and achievement, and the use of technology and achievement in one single study.
Review of related literature

The literature review of this study is presented beneath three subheadings: gender, university level, and achievement; study habits and achievement; and use of technology and achievement.

Gender, university level, and achievement

Conger and Long (2010) examined disadvantages that male students experienced with respect to grade-point average, credits earned, and persistence in college, and found that male students have lower GPAs and fewer credits in their first semester of college largely because they came to college with lower high-school grades. Female students’ better high-school grades explain some of the gender disparity in performance, but differences in college course-taking and majors also explain gender gaps in credits, grades, persistence, and graduation.

A study on persistence and success in an academic program at a community college revealed that students’ GPAs, cumulative hours attempted, and cumulative hours completed were significant predictors of persistence and that young male students were a high-risk group (Stewart & Levin, 2001). Veenstra and Kuyper (2004) also revealed that gender was an important variable for academic achievement and female students were found to have higher levels of motivation. Finally, it was indicated that there was a significant difference between the genders in favor of female students regarding academic achievement in secondary education (Sarier, 2010). The literature did not indicate any finding about the effect of both university level and gender on achievement in one study.

Study habits and achievement

The study habits construct in the current study is twofold: working style (individual or studying in a group) and motivation (intrinsic or extrinsic). Olatoye and Ogunkola (2008) reveal that study habits make a significant contribution to the prediction of achievement in physics. In addition, students who demonstrate higher academic achievement use a variety of study skills compared to students of lower academic achievement (Ergene, 2011; Fazal, Hussain, Majoja & Masood, 2012; Fayombo, 2011; Yu, 2011).

Group learning can contribute to more in-depth learning (Hall, Ramsay, & Raven, 2004). However, some studies indicate that group study is not a significant variable in academic achievement (Yu, 2011). As cited in Yu (2011), Schuman, Walsh, Olson, and Etheridge (1985) examined group studying, cramming, note-taking, reviewing of past exams, and re-reading of material, and concluded that none of these variables had a direct effect on grades.

Motivation, “which is primarily concerned with activation and persistence of behavior, is also partly rooted in cognitive activities” (Bandura, 1977, p. 193). Motivation can influence how, when, and what we learn (Schunk, 1991). There are two types of motivation: intrinsic and extrinsic. The difference between intrinsic and extrinsic motivation is that the former is the doing of an activity because the activity is interesting and enjoyable rather than the instrumental value associated with the activity (Ryan & Deci, 2000). Eymur and Geban (2011) examined the relationship between motivation and academic achievement of pre-service chemistry teachers and found a significant relationship between achievement and two intrinsic motivation subscales (to know and to experience simulation). They also found a significant difference between women and men with respect to intrinsic motivation.

In another study, a positive predictive effect of intrinsic motivation on academic achievement was found for both Indian immigrants and Indian adolescents (Areeppattamannil, Freeman, & Klinger, 2011). Ayub (2010) also found a relationship between intrinsic and extrinsic motivation and the academic performance of college students in India. With regard to gender, the findings reveal that women are more intrinsically motivated, whereas men are more extrinsically motivated. Chan et al. (2012) found that intrinsic motivation had predictive effects on academic achievement for secondary students. The findings of Lynch (2006) also suggest that intrinsic motivation, but not extrinsic motivation, is associated with academic course grades. Conversely, in another study (Liao, Ferdenzi, & Edlin, 2012), motivation was found to not directly affect the academic achievement of either international or domestic students. However, they did find that for international students, both extrinsic and intrinsic motivations indirectly affected academic achievement through the mediating influence of efficacy regarding self-regulated learning, although this was not the case for domestic students.
The studies above had small sample sizes and focused only on either motivation or group learning. The focus of the current study is to examine the effect of working habits (i.e., individual learning and group learning), motivation, and gender on achievement in one single study.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Participants</th>
<th>Method</th>
<th>Variables</th>
<th>Data analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Areepattamannil, Freeman, &amp; Klinger</td>
<td>2011</td>
<td>355 Indian immigrants and 363 Indian adolescents</td>
<td>Descriptive</td>
<td>Intrinsic motivation, extrinsic motivation, and academic achievement</td>
<td>Descriptive, discriminant analysis, hierarchical multiple regression</td>
</tr>
<tr>
<td>Chan, Wong, &amp; Lo</td>
<td>2012</td>
<td>1,381 students</td>
<td>Descriptive</td>
<td>Intrinsic motivation, achievement goals, and learning strategies</td>
<td>Exploratory factor analysis, structural equation modeling</td>
</tr>
<tr>
<td>Eymur &amp; Geban</td>
<td>2011</td>
<td>168 pre-service teachers</td>
<td>Descriptive</td>
<td>Gender, motivation, and achievement</td>
<td>Independent sample t-test, ANOVA, Pearson correlation</td>
</tr>
<tr>
<td>Turner &amp; Croucher</td>
<td>2013</td>
<td>371 undergraduate students</td>
<td>Quantitative</td>
<td>Media-use habits, need for cognition, and grade-point average</td>
<td>Multiple regression analysis</td>
</tr>
<tr>
<td>Sarier</td>
<td>2010</td>
<td>PISA data</td>
<td>Descriptive</td>
<td>Socioeconomic and sociocultural effect and achievement</td>
<td>Descriptive</td>
</tr>
<tr>
<td>Liao, Ferdenzi, &amp; Edlin</td>
<td>2012</td>
<td>310 students</td>
<td>Quantitative</td>
<td>Motivation, self-regulated learning efficacy and academic achievement</td>
<td>ANOVA, linear multivariate regression</td>
</tr>
<tr>
<td>Conger &amp; Long</td>
<td>2010</td>
<td>Not exact number</td>
<td>Quantitative</td>
<td>In grade-point average, credits earned, and persistence in college</td>
<td>Regression and decomposition</td>
</tr>
<tr>
<td>Topcu &amp; Uzundumlu</td>
<td>2012</td>
<td>150 students</td>
<td>Descriptive</td>
<td>student-oriented factors, physiological variables, physical environment,</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>and the variables focused on the academic and the measurement techniques</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>used by them.</td>
<td></td>
</tr>
<tr>
<td>Junco</td>
<td>2012</td>
<td>1839</td>
<td>Descriptive</td>
<td>Time spent on Facebook, frequency checking of Facebook, frequency of</td>
<td>Hierarchical (blocked) linear regression analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Facebook activities, time spent for preparing class</td>
<td></td>
</tr>
<tr>
<td>Whitley</td>
<td>1997</td>
<td>82 studies</td>
<td>Meta-</td>
<td>Gender difference, computer-related attitudes, behavior</td>
<td>Effect size, Q statistics, partial correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hargittai &amp; Hinnant</td>
<td>2010</td>
<td>270 adults</td>
<td>Descriptive</td>
<td>Users’ education, autonomy of use, online experiences, and quality of</td>
<td>Ordinary least squares regression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>equipment</td>
<td></td>
</tr>
</tbody>
</table>

The use of technology and achievement

A new study that reviewed 154 qualifying studies showed that educational technology applications generally had a positive though modest effect in comparison to traditional methods on student achievement in K–12 classrooms (Cheung, 2013). Turner and Corucher (2013) found that college students’ use of traditional media was a significant and viable predictor of both college students’ GPAs and their levels of need for cognition. On the other hand, college students’ use of socially interactive technologies was wholly unrelated to college students’ GPAs and their levels of
need for cognition. Junco (2012) also revealed that there was a strong negative relationship between time spent on social media and academic achievement.

Hargittai and Hinnant (2008) stated that young adults are the most highly connected to the Internet. However, it is not reasonable to say that their Internet usage is homogenous, because young adults in the same age range can use the Internet for a wide range of purposes. Their study show that those with higher levels of education and those with a more resource-rich background use the Web for more “capital enhancing” activities, because they engage in more meaningful activities such as career advancement and political participation. Li and Kirkup (2007) investigated differences in the use of, and attitudes toward, the Internet and computers between Chinese and British students as well as gender differences in this cross-cultural context. Significant gender differences were found in both national groups. For example, male students in both countries were more likely to use email and chat rooms than female students. Volman, Eck, Heemskerk, and Kuiper (2005) investigated the accessibility and attractiveness of different types of ICT applications in education for girls and boys. In this respect, gender differences appeared to be small in primary education. However, the girls’ attitudes toward computers seemed to be less positive than that of boys. In addition, girls and boys engaged in different tasks when working together on computers. Furthermore, with respect to secondary education, the two genders also handled ICT tasks differently. In a meta-analysis of gender differences regarding computer-related attitudes and behaviors, males exhibited greater sex-role stereotyping of computers, higher computer self-efficacy, and more positive affect about computers than did females (Whitley, 1997).

The current study is interested in whether or not students’ routine use of technology is related to achievement. The use of technology is twofold: (a) Internet connection duration and (b) technology use aims. This study examines two constructs and gender in relation to achievement in one single study.

The aim of the study

In sum, the above prior studies indicate the relation of gender, motivation, group learning, and technology usage correlation to academic achievement, but these studies also have some limitations. First of all, they have small sample sizes. Thus, it is hard to generalize such small sample sizes to whole population and use common statistical analyses that require predetermined assumptions such as normality, homogeneity of variance, etc. Secondly, the current study tries to overcome these limitations by working with a large number of learners and by classifying students based on their motivation, working style, technology use aims, and Internet connection duration in the same study. Therefore, this study aims to classify university students’ achievement based on three constructs (demographics, study habits, and motivation and familiarity with technology). The following research questions guided the current study:

- **RQ1**: To what degree do demographics (gender and school year) impact the differences in university students’ GPAs?
- **RQ2**: To what degree do study habits (motivation type and working style) and gender affect the differences in university students’ GPAs?
- **RQ3**: To what degree does technology familiarity (weekly Internet connection duration and number of Internet connection aims) and gender impact the differences in university students’ GPAs?

Method

**Study group and data collection**

Since the aim of the data collection was to reach the most representative sample, this study used a convenience sampling method as heterogeneous samples include a wide variety of characteristics to which the results of the study may be generalizable. The researchers sought out various schools to obtain a heterogeneous sample (Fraenkel & Wallen, 2006; Heppner, Wampold, & Kivlighan, 2008). For this study, random sampling, stratified sampling, or a combination of the two were not used as sampling methods for several reasons. First, some course instructors did not deliver surveys to their students due to their time constraints. Second, some students were not willing to participate in the study. Third, it was difficult to locate some courses or schedule time to administer the survey, etc. In the end, the data were obtained from 2,949 respondents from 11 faculties of a public university in Turkey. The participation ratio from the faculties was 25.97% (n = 766) from the faculty of education, 21.26% (n = 627) from the faculty of
engineering, 14.1% (n = 415) from the faculty of economics, and administrative sciences, 7.62% (n = 225) from the faculty of law, 6.47% (n = 191) from the faculty of literature, 6.27% (n = 185) from the faculty of medicine, 4.64% (n = 137) from the faculty of business, 4.10% (n = 121) from the faculty of science, 3.86% (n = 114) from faculty of fine arts, 3.25% (n = 96) from the maritime faculty and 2.44% (n = 72) from the school of nursing.

Instrumentation and manipulation of variables

The data-collection tool is a survey that includes questions about respondents’ self-reported GPAs, demographic information, study habits, and familiarity with technology. The GPA is the hypothesized construct and the dependent variable of the study. The dependent variable should exactly reflect the construct to which it may be affected by the independent variables. Accordingly, the researchers asked the students for their cumulative GPAs to determine their achievement in higher education. Self-reported GPAs have also been used by other researchers, and these researchers have reported that official GPAs are correlated with self-reported GPAs (Kuncel, Crede, & Thomas, 2005; Schuman, Walsh, Olson, and Etheridge, 2013).

The independent variables, as determined by the researchers, focused on three main variables of interest, demographics (gender and school year), study habits (working style, motivation types) and technology familiarity (amount of time spent on the Internet, and reasons for Internet use). First, we ascertained gender and university year of respondents were ascertained. Second, the students were asked about their working style (individual or group learning). They were also asked about their motivation to study (extrovert or introvert). Third, the students were asked about their amount of time spent of the Internet during an average week (0, 1–7 hours, 8–21 hours, 22–35 hours, or more than 36 hours) and reasons for using the Internet (studying, shopping, interaction, doing homework, online banking, social sharing, listening to radio, watching TV, reading newspapers, gaming, Internet surfing, listening to music, watching movies, and other reasons).

Before analyzing the data, we manipulated the independent variables to determine whether a significant difference among the independent variables existed (Heppner at al., 2008). In this study, all independent variables were identified as categorical variables. The university year was limited to four categories (first, second, third, and fourth and above) because the number of the students whose school year exceeded four did not reach the expected ratio. Similarly, the number of hours students spent on the Internet each week was also combined into four categories because there were too few students who never connected to the Internet during the week. This scale was then converted to a categorical variable: 0–3, 4–6, and 7–11. Table 2 shows the number and percentage of the students based on the manipulated levels of independent variables.

<table>
<thead>
<tr>
<th>Table 2. The participants’ demographics, study habits, and familiarity with technology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent variables</strong></td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
</tr>
<tr>
<td>School year</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Study habits</strong></td>
</tr>
<tr>
<td>Working style</td>
</tr>
<tr>
<td>Motivation type</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Technology familiarization</strong></td>
</tr>
<tr>
<td>Amount of time spent on the Internet each week</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Internet connection aims</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Data analysis

The CHAID algorithm reveals a tree including meaningful nodes that classify a nominal or ordinal dependent variable (Magidson & Vermunt, 2005). The first CHAID algorithm uses Chi-squire goodness of fit test to identify significant branching regions of independent variables and then merges some levels of independent variables that do not differ in the prediction of the dependent variable (Magidson & Vermunt, 2005). Departing from the classification and regression tree (CART), the CHAID algorithm allows for more than two branching, if there exist significant differences. If the dependent variable is ordinal, the CHAID algorithm uses F statistics to decide whether there is another significant predictor of the dependent variable. In the current study, the researchers used ordinal GPA scores and $F$ statistics to produce the tree.

When comparing with other inferential statistics methods, the CHAID algorithm has specific advantages. First, while inferential statistical analysis requires the control of several assumptions for statistical tests (Green & Salkind, 2008), the CHAID algorithm does not. Second, the CHAID algorithm makes decisions about dependent variables owing to several terminal nodes in the tree model, while inferential statistics evaluates only whether there exists a significant difference among mean scores of dependent variables in each category of independent variables. Therefore, in this study, we preferred to use the CHAID algorithm to obtain a decision tree that indicated the most significant breaks in the prediction of academic achievement of university students, and accordingly, we aimed to reveal the characteristics of successful and unsuccessful students who are pursuing their higher education.

In addition, we used logistic regression to determine whether the results of the CHAID algorithm were compatible with the results of the logistic regression. Logistic regression aims to predict group membership from a set of independent variables that may be continuous, discrete, dichotomous, or a combination thereof (Tabachnick & Fidel, 2001, p. 517). Logistic regression does not require independent and dependent variables to be continuous variables. In addition, a normality assumption is not necessary to conduct an analysis (Cokluk, Sekercioglu, & Buyukozturk, 2012). In this study, the independent variables are dichotomous and categorical, and the use of logistic regression aims to predict successful university students from a set of their demographics, study habits and technology familiarity.

We determined a dichotomous dependent variable of logistic regression by grouping the first 25% quartile of the students based on their GPAs in one group and coding this group as group 1. The GPAs of the students in group 1 ranged from the lowest grade to a 2.07. Respondents in group 1 could be considered the unsuccessful students. The other respondents were coded as group 2, which means that the GPAs of the students in this group ranged from 2.0701 to the highest grade.

Findings

Classification of GPAs based on gender and year of university

The CHAID algorithm created a tree model with branches for GPAs based on gender and school year. In the tree model, respondents’ GPAs were divided by regions for the independent variables and yielded four main results. The first result of the tree reveals that female respondents ($\bar{X} = 2.62; SD = 0.58; n = 1439$) are significantly more successful than male respondents ($\bar{X} = 2.33; SD = 0.56; n = 1180$). Second, male respondents in the first year are likely to have the lowest GPAs among all respondents ($\bar{X} = 2.13; SD = 0.72; n = 254$). Similarly, female respondents in first year are likely to have the lowest GPAs among the female respondents. These two results indicate that, at the university, first-year female and male students are more unsuccessful than students attending higher years. The third result suggests that male respondents in the third year are likely to have the highest GPAs among the male respondents. Fourth, female respondents in fourth year and above are likely to have the highest GPAs among all the respondents. These findings are consistent with the other studies that reveal GPAs differ based on gender (Conger & Long, 2010; Turner & Croucher, 2013).
Classification of GPAs based on gender and study habits

The second tree model indicates a branching out for GPAs based on motivation type, gender, and working style. The group with the lowest GPAs is the male group, who demonstrate extrovert motivation and prefer learning in a group ($\bar{X} = 2.21; SD = 0.48; n = 207$). Next, the group with the highest GPAs is the female group who demonstrate introvert motivation ($\bar{X} = 2.67; SD = 0.57; n = 757$). The most valuable result from this tree is that, regardless of gender, those students classified as introvert have higher GPAs than those classified as extrovert. Furthermore, respondents who prefer to study individually have higher GPAs than those who prefer to study in a group.
Classification of GPAs based on gender and technology familiarity

The third tree model shows a branching out on GPAs based on the amount of time spent on the Internet each week, the number of reasons for using the Internet, and gender. The tree shows that male students using the Internet more than 22 hours a week and using the Internet for up to six reasons have the lowest GPAs among all respondents ($\bar{X} = 2.134; SD = 0.63; n = 123$). On the other hand, female respondents using the Internet up to 21 hours a week have the highest GPAs among all respondents ($\bar{X} = 2.63; SD = 0.57; n = 1092$). These two results are important, because they suggest that if respondents’ use the Internet more than 22 hours a week, their GPAs may decrease.

**Figure 3. Classification of GPAs based on gender and familiarity with technology**

Comparison of CHAID with logistic regression results

Block 1 includes all independent variables and constants that predict whether a student is in the 25th percentile. The prediction is based on the independent variables, which show a total correct classification ratio for the model to be 75.2%. The omnibus test of model coefficients are significant ($\chi^2 = 189.194; p < 0.05$). The independent variables produce a significantly improved prediction of whether a student is unsuccessful (Acton, Miller, Maltby, & Fullerton, 2009; Cokluk et al., 2012). A Hosmer and Leme test (for goodness of fit) indicated that the model is fit for the data ($\chi^2 = 11.35; p = 0.18$). When independent variables enter the analysis, an 8% (according to Cox & Snell $R^2$) or 11% (according to Nagelkerke $R^2$) variation in academic performance is explained by the independent variables. Table 3 indicates that demographics, study habits, and familiarity with technology have a significant effect on student performance at the university level.

With regard to demographics, the results indicate that male students are 2.33 times more likely to be in the 25th percentile than female students. Similarly, students in their first year are more likely to be in the 25th percentile than second-, third-, and fourth-year students. In addition, a student using the Internet 0–7 hours a week will be approximately 1.5 times more likely to be in the 25th percentile than a student using the Internet more than 35 hours. With regard to motivation, extrovert students will be 1.35 times more likely to be in the 25th percentile than introvert students.
Similar to the CHAID analysis, the logistic regression finds that first-year extrovert male students are more unsuccessful than all other students. This comparison indicates that the CHAID analysis achieves greater classification than other methods because this analysis includes more than one independent variable as well as the levels of these variables in the same analysis.

**Table 3. Logistic regression results**

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School year (first-year is ref. category)</td>
<td>2</td>
<td>.819</td>
<td>71.644</td>
<td>3</td>
<td>.000</td>
<td>2.269</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.001</td>
<td>10.179</td>
<td>1</td>
<td>.000</td>
<td>2.720</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.867</td>
<td>6.709</td>
<td>1</td>
<td>.000</td>
<td>2.380</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>.844</td>
<td>.098</td>
<td>17.408</td>
<td>1</td>
<td>.000</td>
<td>2.326</td>
</tr>
<tr>
<td><strong>Technology familiarity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly Internet use (0–7 is ref. category)</td>
<td>8–21</td>
<td>.056</td>
<td>8.146</td>
<td>1</td>
<td>.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22–35</td>
<td>.208</td>
<td>11.370</td>
<td>1</td>
<td>.152</td>
<td></td>
</tr>
<tr>
<td>more than 35</td>
<td>.410</td>
<td>.146</td>
<td>7.836</td>
<td>1</td>
<td>.005</td>
<td>1.507</td>
</tr>
<tr>
<td>Internet use aim (1–3 is ref. category)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4–6</td>
<td>.198</td>
<td>4.670</td>
<td>1</td>
<td>.253</td>
<td>1.219</td>
</tr>
<tr>
<td></td>
<td>7–11</td>
<td>.269</td>
<td>4.211</td>
<td>1</td>
<td>.120</td>
<td>1.309</td>
</tr>
<tr>
<td><strong>Study habits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d13 (group) (individual is ref. category)</td>
<td>.131</td>
<td>.102</td>
<td>1.651</td>
<td>1</td>
<td>.199</td>
<td>1.140</td>
</tr>
<tr>
<td>d15 (extrovert) (introvert is ref. category)</td>
<td>.301</td>
<td>.096</td>
<td>9.904</td>
<td>1</td>
<td>.002</td>
<td>1.351</td>
</tr>
<tr>
<td>Constant</td>
<td>-.831</td>
<td>.233</td>
<td>12.739</td>
<td>1</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

**Discussion**

This study reveals relationships involving university students’ demographics, study habits, and familiarity with technology are correlated with their self-reported GPAs. Different from available studies, this study indicates that inferences can be made based on several terminal nodes taken from the tree model, thus making it easier to identify significant breaks that involve one or more independent variables. The results of this study show that gender, study habits, and familiarity with technology are important factors that may explain university students’ achievement.

Consistent with the findings of similar studies, the results of this study also indicate that female students are more successful than male students (Conger & Long, 2010; Veenstra & Kuyper, 2004; Sarier, 2010). Conger and Long (2010) revealed that male students earned lower GPAs and fewer credits in their first semester of college, largely because they arrived with lower high-school grades, while female students’ better high-school grades explained some of the gender disparity in performance. Different from available studies, the findings of the current study further show that male students in the first year are likely to have the lowest GPAs among all respondents. This result suggests that male students are at greater risk of being unsuccessful academically than female students because they are behind their female counterparts in each school year. This result is also consistent with the extant literature (Stewart & Levin, 2001). Similarly, female students in first year are likely to have the lowest GPAs among all female respondents. That is, the current study indicates that both male and female students may earn lower GPAs in the first academic year because it is their first encounter with the higher education. In Turkey, prior to attending higher education, students graduate from general high schools or vocational/technical high schools (Aksit, 2007). Only students who receive the best scores on the university entrance exams can then enroll in schools of medicine, dentistry, political science, etc. This suggests that new incoming students do not have time to adapt to university life after leaving high school. The literature illustrates that the reason for the lower GPAs in first year might be due to several reasons, such as communication problems that the students encounter, differences in teaching methodologies and learning strategies at the university level, the physical environment, and a lack of awareness regarding their personal study skills (Topcu & Uzundumlu, 2012). Furthermore, the findings indicate that, for both male and female students, GPAs increase as their progress with each year of university.
This classification study shows that female students who are introvertedly motivated have the highest GPAs, while males who are extrovertedly motivated having the lowest GPAs. Consistent with the existing literature, Ayub (2010) also found that female students were more intrinsically motivated, while male students were more extrinsically motivated. Regardless of gender, however, students who are classified as an introvert have higher GPAs than those who are found to be an extrovert. Based on previous research, it can be concluded that intrinsic motivation has a positive effect on achievement (Areepattamannil et al., 2011; Eymur & Geban, 2012; Lynch, 2006). In terms of working habits (individual or group learning), students who prefer to study individually have higher GPAs than their counterparts who prefer group study. Therefore, it is important to examine our students’ group processing as students engage with their peers while learning and studying in a group (Johnson & Johnson, 2009). With respect to learning groups, if the teacher or researcher ensures that every group member works actively in the construction of knowledge, scaffolds other group members, and works to the best of their ability, the performance and achievement of the group and of each member may increase (Goodsell, Maher, & Tinto, 1992; Vygotsky, 1978). Conversely, if there are deficits in the group work or if group members do not know how to communicate their knowledge to others, the group as a whole and the individual members of the group cannot attain their desired goals.

Given the above findings, we recommend that well-designed group work (which is very important for higher education because approximately one-third of the students prefer to learn in a group) be incorporated in the curricula. Furthermore, the use of small group design using technology should be supported by higher education institutions because there is considerable positive published evidence that supports its implementation (Bennett, Bishop, Dalgarno, Waycott, & Kennedy, 2012; Laru, Naykki, & Jarvela, 2012; Li, Helou, & Gillet, 2012). Apart from these available studies, which focus on either motivation or group learning, the current study examines motivation type, group learning, and gender in one study. The findings indicate that achievement of extrovert male students preferring learning in a group is lower than introvert male students preferring learning in a group. The similar result is not meaningful for female students.

Finally, the results illustrate that Internet usage totaling more than 22 hours a week may result in decreased GPAs and that technology may have a detrimental effect on school work. Studies on the effect of ICT on academic performance provide more evidence that a high amount of technology use causes some problems for students, such as lower achievement and Internet addiction (Chou & Hsiao, 2000; Chou, Condron, & Belland, 2005; Junco, 2012). While net-generation students are capable of increased multitasking, this multitasking may cause academic delays of which they are not aware (Bowman, Levine, Waite, & Gendron, 2010; Junco & Cotten, 2012). Therefore, it is vital to inform students about how to use the Internet effectively and in such a way that it contributes to their success in higher education. The current study, however, reveals that Internet use of up to 21 hours, which appears to be a critical point, can explain university students’ higher GPAs, as university students tend to use the Internet for academic purposes up to this point. Increasing services available on the Internet, such as academic libraries, online certifications, and university communities on social media, allow students to spend more time on the Internet (Hendrix, Chiarella, Hasman, Murphy, & Zaphron, 2009). Most of our sample includes Net generation students and, as such, they have higher level computer skills and more experience with social media than did students 10 years ago (Frand, 2000; Oblinger & Oblinger, 2005). As a result of these factors, higher education curricula should incorporate the effective use of technology, as also emphasized by another study and instructors should adapt their lessons according to the characteristics of students of the Net Generation.

The researchers of this study also suggest that other researchers should focus on more pedagogical aspect of this topic in the future. We certainly think and believe that other researchers should study professions such as engineering or teacher education. Therefore, in addition to gender and university level, university profession should be added to the study. With regard to study habits, self-efficacy, students’ satisfaction with the education, and their interest to the profession may be correlated with their achievement. Lastly, an ICT literacy construct, which may include “information technology, mobile technology, and communication technology literacies,” can be added to the study. This construct will let researchers to see which technology we should select while designing a course in university level.

References


Exploring Elementary-School Students’ Engagement Patterns in a Game-Based Learning Environment

Ya-Hui Hsieh, Yi-Chun Lin and Huei-Tse Hou*

National Taiwan University of Science and Technology, Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, #43 Keelung Road, Section 4, Taipei, Taiwan // only_no1@hotmail.com // hthou@mail.ntust.edu.tw

*Corresponding author

(Submitted January 21, 2014; Revised August 7, 2014; Accepted September 3, 2014)

ABSTRACT

Unlike most research, which has primarily examined the players’ interest in or attitude toward game-based learning through questionnaires, the purpose of this empirical study is to explore students’ engagement patterns by qualitative observation and sequential analysis to visualize and better understand their game-based learning process. We studied the sequential behaviors of 34 students (17 male and 17 female) and considered issues of gender differences and sequential pattern similarities. The results show that the behavioral coding schema provided by the authors and the innovative method of sequential analysis can provide researchers with a certain level of understanding of students’ engagement patterns in game-based learning environments. In terms of the overall sequence results, this study identified higher and lower engagement patterns to represent students’ learning processes in game-based learning. Moreover, the sequential patterns represent qualitative differences and similarities in engagement patterns grouped by gender. A good engagement cycle, in which male and female students started the game to attempt to think and solve problems, was noticeable. However, male students were observed to demonstrate more engaged behaviors, with continuous self-conversations when confused. The frequency of self-conversation from female students was obviously lower than that of male students and revealed more verbal and nonverbal behaviors. The deep examination of students’ verbal and nonverbal engagement behaviors may make beneficial contributions to the educational technology field with the adoption of sequential analysis.

Keywords

Engagement pattern, Sequential behavioral pattern, Gender difference, Game-based learning

Introduction

In recent years, game-based learning has become a critical issue in education (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012) because games may help students engage with learning by offering fantasy and academic content (Chen, 2014; Chang, Wu, Weng, & Sung, 2012). According to previous studies, students’ involvement and participation in learning activities should be constructed based on positive interactions between students and their learning environment (O’Brien & Toms, 2008). To fulfill this requirement, a game must embed fantasy content to motivate students to active participate in the task (Gunter et al., 2008). Moreover, a game should provide sufficient challenges, immediate and clear feedback, and playable experiences to increase students’ engagement in the learning process (Inal & Cagiltay, 2007; Hou & Li, 2014; Kiili, 2005).

The game design mechanism includes the game goals and the learning goals (Gunter, Kenny, & Vick, 2008; Ke & Abras, 2012). The game goals must be designed to attract students’ attention and to encourage their engagement in game-based learning activities (Alessi & Trollip, 2001). Therefore, we can optimistically speculate that students acquire knowledge through the game play process to achieve learning goals. Based on Chen (2014), the mini-game approach benefits the development of GBL in two ways. First, the mini-game approach maintains the structure of the materials, which makes it suitable for well-structured subject domains, where the learning materials can be decomposed into smaller lessons or units to be drilled upon. Second, the mini-game approach offers more flexibility for application in domain-independent subjects. That is, the mini-game is allowed to utilize a signal principle to set knowledge or skills that students are required to learn (Maertens, Vandewaetere, Cornillie, & Desmet, 2014). However, the influences of mini-game-based learning on students’ engagement behaviors are still unclear.
To confirm whether game-based learning effectively facilitates students’ learning, some available methods, such as interviews, observations, surveys, and usability tests, have been used (Reeves & Hedberg, 2003). For example, based on the technological acceptance model, researchers suggest that students’ experience of both usefulness and ease of use of game design features are related to their attitude toward game-based learning (Hou & Li, 2014). In addition, researchers have emphasized the relationships among game interfaces, users’ learning-skill development, and achievement (Pinelle, Wong, & Stach, 2008).

Generally, while learning in a game-based environment, individuals must engage in activities such as problem identification, hypothesis making, and reflective thinking (Maertens et al., 2014). The degree of engagement is positively related to the individual learning outcomes (Admiraal, Huijzenga, Akkerman, & Dam, 2011). Typically, researchers have proposed a three-part typology, emphasizing affective, behavioral, and cognitive dimensions of engagement (Fredericks et al., 2004). Behavioral engagement refers to the observable behaviors reflected in attendance or active participation, such as asking questions or participating in discussions. Cognitive engagement refers to an individual’s goal setting, self-regulation of performance, or application of learning strategies. Affective engagement refers to attitude — a sense of intention, interest, or motivation to engage in a task.

Most research has examined students’ engagement behaviors in game-based learning through questionnaires. For example, most researchers have used flow assessments to measure engagement (Admiraal et al., 2011; Hsieh, Lin, & Hou, 2013; Hou & Li, 2014; Inal & Cagiltay, 2007; Kiili, 2005). However, standard questionnaires may not be easily incorporated into various game-based learning environments. Moreover, important information that arises during the task execution process may be lost. For example, the elements of the game may produce physical reactions (Barsalou, 2003); at the same time, every action is associated with an instant reward or feedback from the game (Ke & Abras, 2012). Thus, unlike most research, which has examined only the students’ attitude or final performance of game-based learning through questionnaires, qualitative observations of students’ behavioral engagement with game-based learning are fundamental for the educational technology field.

Interaction with the digital application may also increase users’ cognitive and affective experience (Weibel, Wissmath, Habegger, Steiner, & Groner, 2008), and some of these experiences will be embodied by physical behaviors (Smith, Cottrell, Gosselin, & Schyns, 2005). For example, a previous study found that students showed verbal behaviors, such as murmuring, expressing strongly positive/negative comments, asking questions to themselves, or stating their confusion or frustration while navigating the digital application (Tullis & Albert, 2008). The private speech may play a critical role in individual’s thinking and action. Private speech is critical for the development of self-regulation and helps individuals decide whether they acted the way that they had planned. Thus, observing verbal information may help educators understand students’ inward reflection on what they are doing and how these behaviors guide their own actions purposefully.

In an interactive context, observation of reactive behavior was a more reliable indicator than children’s responses to questions (Hanna, Risden, & Alexander, 1997). Some nonverbal behaviors were observed, including facial expressions and body language directed toward the digital application task (Tullis & Albert, 2008). Sometimes, students with multiple learning disabilities are often unable to talk to teachers. Such important nonverbal behaviors may provide preliminary clues to help researchers understand students’ learning experience and evaluate the appropriateness of the design of game mechanism. However, rarely studies explore the effects of these potential behaviors. To know more about what type of verbal and nonverbal behaviors occur in a game-based learning process may inform us regarding the relationships between students and the design of the game. Therefore, to understand the effectiveness of game-based learning, we believe that the scope of the methodology should be extended to better understand users’ reaction.

Gender differences in game-based learning have recently received increasing attention from researchers. Although gender difference is not a new issue, gender sensitivity toward educational games is strongly needed (Robertson, 2012). However, a large body of empirical research has noted the absence of gender differences in learning performance and motivation (Ke & Grabowski, 2012; Papastergiou, 2009). For example, to investigate whether there are gender differences in the game-making skills displayed by students, Robertson (2012) found that female students chose to spend more time writing dialogue than did boys, and female students scored higher in games than did male students. Hou and Li (2014) evaluated multiple aspects of problem-solving-based educational games and found that male students reported more positive results than did female students for the flow antecedents of challenge, goals, feedback, and playability. However, several studies found no significant gender differences in learning and
motivation in game-based learning (Ke & Grabowski, 2012; Papastergiou, 2009). Accordingly, the impact of gender on game-based learning may reveal different results based on different game mechanisms. Additionally, because qualitative empirical research on gender issues in game-based learning remains limited (Papastergiou, 2009), it is necessary to explore gender differences in learning patterns with game-based learning.

In sum, this study suggests that a game design mechanism that matches the features of the game’s goal may produce corresponding behavior patterns. To a certain extent, these behavior patterns may represent students’ learning states, such as engagement behaviors, which are essential in achieving learning goals. Observation may be a highly appropriate method to collect data about students’ dynamic behaviors in game-based learning environments. Understanding students’ behavior patterns may help us to examine the relationship between the learning and gaming processes. Careful collection of students’ verbal and nonverbal behaviors may be helpful in discovering the way that individual behaviors contribute to these processes in a game-based learning environment. Recently, sequential analysis has been integrated into behavioral analysis methods to examine the overall learning process (Hou, 2012a). Sequential analysis allows researchers to identify the sequential relationships of behaviors (Bakeman & Gottman, 1997). Through the visualized results of sequential analyses, we can see how learning occurs for students from inside out. Additionally, research has utilized sequential analysis method to analyze the behavioral patterns in massively multiple online role-playing games (Hou, 2012b).

Therefore, this study moves away from normative evaluations of students’ achievement and focuses on the game-based learning process. This empirical study focuses on students’ game-based learning process using a sequential analysis. The main purpose of this study is to explore students’ engagement patterns in the game-based learning environment through integrating observation, content analysis, and sequential analysis. The questions to be addressed in this study are as follows:
• In a game-based learning environment, what is the distribution of the engagement behaviors demonstrated by students?
• What are the overall sequential engagement patterns displayed by students?
• What are the differences and similarities in behavior patterns between genders?

Method

Participants

The participants included 34 fourth- to sixth-grade elementary students (17 males and 17 females). Using a purposive sampling method, these students were selected from a cram school in Taiwan. All of these students attended the cram school for additional instruction in English, Chinese, and mathematics. All of them had experience in online games; their average time spent playing computer games was five to ten hours per week.

Game introduction

In Taiwan, students from elementary school to college are asked to perform resource classification. Resource classification is made based on their material and recyclable value. According to the current promotion of the government, resource classification is a public movement in train stations, airports, movie theaters, and streets. The resource classification ability has become necessary for all people in Taiwan.

Accordingly, a resource classification matching game called “Happy Black-faced Spoonbill” (Lin, Hsieh, Hou, Yen, Chou, & Chen, 2011) served as the platform for game-based learning in this study. The game was designed and developed using the mini-game approach to teach resource classification concepts to elementary school students. The mini-game is short and designed based on a signal principle. Based on the theory of experiential learning (Kolb, 1984), knowledge is gained through both personal and environmental experience during game play process. Experiential learning focuses on developing concrete experience and reflective observation, abstract conceptualization, and active experimentation (Kolb, 1984). Thus, through its rules, the game allows students to engage in direct experience and reflect on their learning. As students go through the experience of playing this game over and over, they learn to develop their own ideas, try them out, and generate concrete concepts based on their
experiences. While navigating within the game environment, each student has the opportunity to think about the resource classification concepts and to test their understanding of these concepts (Lin et al., 2011).

As shown in Figure 1, the game includes a cover story, clear rules and missions, a main screen, and the ending story, in that order. In the game, students must correctly find at least three types of resources to complete the game’s mission. Students must execute the game task individually within a restricted timeframe (three minutes per run) and with ten health points. The time pressure and limited health points are regarded as one of the challenges of the game (Lin et al., 2011). A marker is shown under each recycle bin for students to evaluate their progress. Feedback is provided on the bottom right corner of the screen, with an angry black-faced spoonbill face if the player classifies a resource in an incorrect recycle bin. When students collect more than three resource objects for each classification, they are given a badge as a reward in the upper right side of the screen. When the game is over, score statistics are provided for the performance of the entire play session. Finally, the ending story is shown to promote the important concept of resource classification.

**Figure 1.** Interface of the “happy black-faced spoonbill” (A refers to the health points; B to the main screen for playing; C to recycle bins; D to score statistics; E to the countdown timer; F to badges as rewards; G to the angry face as feedback for wrong answers)

**Procedure**

This study was conducted in a cram school, an after-hours tutoring program. First, to control the effect of students’ prior knowledge about resource classification, each student received individual, informal instruction for less than five minutes about basic concepts of resource classification. Then, the students were invited to play the game individually. Except for the cover story, introduction to the rules, and ending story, each run of the game play takes three minutes to execute. Each student was allowed to play the game repeatedly. After completing the game play, the students were given a brief review to increase their familiarity with the classification conceptions or to clarify their misconceptions.

**Data collection**

A synchronous video-capturing system and web camera were used to collect students’ engagement behaviors. For example, the video-capturing system automatically recorded the computer screen and game background. Simultaneously, the web camera collected students’ voices, facial expressions, and posture to understand how students reacted to the game mechanism. Verbal behavior refers to what students actually say, and the nonverbal behavior refers to students’ facial expressions and body language. In this study, we chose the first barricade for further analysis and discussion. In all, the students contributed 1,899 behaviors during the game-playing process, and each behavior was categorized by content analysis.
Data analysis

Two stages of analysis were executed on the data analysis in this study: (1) content analysis of individual students’ video data and (2) sequential analysis to assess students’ engagement patterns.

Content analysis

Process videotapes were reviewed and coded. Using content analysis, we analyzed students’ video-recording data for behavior distribution. We adopted a coding schema from Tullis and Albert (2008) and Hanna et al. (1997) as the reference to assess students’ reactive behaviors to the digital application. The schema reflected the extent of students’ general active engagement in a learning activity. Through the open coding and constant comparison of data, researchers generated the core categories (shown in Table 1). Episodes of student behavior related to students’ verbal and nonverbal behavior were transcribed.

Each student had three minutes of video recording data. Based on muscular evaluation, previous research has indicated strong facial reactions 500 to 1000 milliseconds after the stimulus presentation (Moody, McIntosh, Mann, & Weisser, 2007). However, because this study attempted to artificially observe verbal and nonverbal behaviors in video data, five seconds was considered an acceptable coding block. If a block showed two or more different behaviors, coding was performed based on the order of appearance. Based on the coding framework, if the block did not show obvious behaviors related to the categories of the table, “OT” was coded. After observing every behavior from all students, two senior researchers coded each behavior based on the schema. The coders categorized all of the behaviors and discussed them to reach consensus. The inter-rater Kappa for these analyses was greater than 0.82 (p < 0.001).

Table 1. Coding schema of verbal and nonverbal behaviors

<table>
<thead>
<tr>
<th>Categories</th>
<th>Code</th>
<th>Description of students’ behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal — Self-question</td>
<td>VS</td>
<td>Student asks himself/herself questions</td>
</tr>
<tr>
<td>Verbal — Frustration</td>
<td>VF</td>
<td>Student expresses frustration</td>
</tr>
<tr>
<td>Verbal — Murmur</td>
<td>VM</td>
<td>Student murmurs to himself/herself</td>
</tr>
<tr>
<td>Nonverbal behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smile</td>
<td>SM</td>
<td>Smiling and a noticeable up-twist of student’s lips</td>
</tr>
<tr>
<td>Focus</td>
<td>FO</td>
<td>Focusing on the screen and seldom moving the head and body</td>
</tr>
<tr>
<td>Close</td>
<td>CL</td>
<td>Moving closer to the screen</td>
</tr>
<tr>
<td>Leave</td>
<td>LE</td>
<td>Moving body away from the screen</td>
</tr>
<tr>
<td>Scratch</td>
<td>SC</td>
<td>Scratching face, hair, or body</td>
</tr>
<tr>
<td>Arrow path</td>
<td>AR</td>
<td>Student’s eyes and head move following the arrow path</td>
</tr>
<tr>
<td>Other behaviors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>OT</td>
<td>Behaviors not included in the above, consisting mostly of unapparent behaviors, such as sitting motionless. These behaviors are difficult to categorize, and it is difficult to identify whether students have dropped out of the game or are attempting to find the solution.</td>
</tr>
</tbody>
</table>

Sequential analysis

To investigate a possible connection between students’ verbal and nonverbal behaviors, we used a sequential analysis proposed by Bakeman and Gottman (1997) to discern the behavioral patterns in this study. The sequential analysis was used to visualize sequential correlations between chronologically ordered behaviors (Bakeman & Gottman, 1997; Hou, 2012a). To meet the requirements of the sequential analysis, students’ behaviors were coded in the chronological order of their occurrence. For example, after focusing on the main task (FO), a student expressed his/her frustration (VF), followed by a smile (SM), and he/she may have been close to the screen (CL). This coding was described as FO→VF, VF→SM, SM→CL. Then, we calculated the frequency transfer matrix, the condition probability matrix, and the expected-value matrix. The calculated Z-scores from the above three matrixes were used to examine the significance of each sequence. Finally, we consolidated significant sequences into a sequential graphic summary to understand the sequential correlations among each behavior.
Result and discussion

Descriptive findings of the content analysis

As shown in Table 2, nonverbal behaviors were found to be the major distribution (94%). Within the total set of nonverbal behaviors, FO (focus on the screen and seldom moving the head and body) was identified at a relatively high frequency (44%). Moreover, the students were frequently found close to the screen (CL) (moving the body forward and close to the screen). A number of behaviors were identified, such as AR (eyes following the arrow path) (7%), SM (smiling) (4%), SC (scratching head or body) (2%), and LE (leaving the screen) (2%). The descriptive results are in line with the suggestion that games should involve students in an activity (Killi, 2005) and that long-term focus should be part of the learning task (Alessi & Trollip, 2001; O’Brien & Toms, 2008). For example, we found that students frequently moved their body to get closer to the screen and seldom took their eyes away from the screen. Moreover, after interacting with the game mechanism, students scratched their heads or smiled. These behaviors demonstrated that the game in this study stimulates students’ curiosity and allowed students to experience pleasure (Alessi & Trollip, 2001; O’Brien & Toms, 2008).

In contrast to nonverbal behaviors, verbal behaviors (6%) were observed less frequently, such as VM (murmuring) (3%), VS (self-questioning) (1%), and VF (expressing frustration) (2%). For example, behaviors such as saying, “Oh my God, how does this work?” “How come? What happened?” or “Oops, wrong again” were observed. The verbal behaviors in this study provide evidence of an engaging experience that was aroused by the challenge of the game. Researchers indicated that the challenge experience may prompt students to reflect on their thinking and to revise what they have done (Woszczynski, Roth, & Segars, 2002). We suggest that verbal behaviors may have an implicit influence and can be regulated during students’ gameplay process.

Additionally, gender differences have been found in verbal and nonverbal behaviors. Male students expressed more VS (self-questioning), VF (expressing frustration), and VM (murmuring). In addition, male students tended to demonstrate SM (smiling), CL (moving closer to the screen), LE (moving body away from the screen), and AR (eyes following the arrow path). In contrast, female students behaved with more FO (focusing on the screen and seldom moving the head and body) and SC (scratching face, hair, or body) behaviors. The distribution of the behaviors (OT) that were not related to the category schema was 5%. However, the verbal, nonverbal, and OT behaviors do not exist independently. Because the engagement assertion is not substantiated by the separate results, we present the results of the sequential analysis.

<table>
<thead>
<tr>
<th>Table 2. Frequency distribution of students’ engagement behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal behavior</td>
</tr>
<tr>
<td>VS</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Male (n = 17)</td>
</tr>
<tr>
<td>Female (n = 17)</td>
</tr>
<tr>
<td>All (n = 34)</td>
</tr>
</tbody>
</table>

Results of the sequential analysis

The sequential analysis helped us to identify the significant sequence of behavior transference during the game play process. The results are shown in the tables and figures. In all tables, the rows indicate the starting behaviors, and the columns indicate the follow-up behaviors. The Z-scores of the table were greater than 1.96, indicating that behavioral continuity from a certain row to a certain column reached statistical significance (p < 0.05). According to the significant sequence, we further inferred a sequential engagement pattern diagram, as illustrated in the figures. In the figures, the arrows indicate the direction of the significant behavior sequence, and the thickness of the arrowhead line indicates the level of significance. The circle represents students’ nonverbal information (facial expressions and changes in body language), verbal information, and OT behaviors.
Table 3 and Figure 2 present students’ overall engagement patterns while learning in the game-based learning environment. Reciprocal relationships among verbal, nonverbal, and OT behaviors were found. For example, we observed that students expressed smiling or scratching behaviors after expressing frustration (VF→SM, VF→SC), self-questioning (VS→SM), and murmuring (VM→SM). Often, after smiling behavior, students returned to self-questioning (SM→VS), expressed frustration (SM→VF), or murmured (SM→VM). For example, some students murmured with a smile, “Oops, what happened to me?” or “Well, perhaps the battery belongs to this.” Moreover, the subsequent scratching behavior may indicate students’ confusion (SM→SC, SC→VS, SC→VF). Blank moments prior to other actions or actions progressing to blank moments were also observed. For example, we observed that students went into a blank moment approximately 5 to 15 seconds after asking themselves questions or murmuring (VS→0, VM→0), followed by self-questioning and murmuring (behavior and continuous attention to the game task. For example, while executing the task, students focused on the screen thoroughly and then moved their bodies closer to the screen (FO→CL; CL→FO). Sometimes students’ bodies were close to the screen to reduce their distance from the avatar in the game, with their eyes following the arrow path to focus on the game elements (CL→AR, AR→FO). This pattern was repeated, and students were observed contemplating during the silent period. The reciprocal transference from the nonverbal to nonverbal behaviors seemed to be related to researchers’ suggestions that games should evoke students’ concentration, capture students’ attention, and propel students toward engagement, leading to subsequent learning behaviors (Alessi & Trollip, 2001). Accordingly, lower engagement patterns may help in assessing whether a game is accessible and attainable.

Table 3. Adjusted residuals table of overall students (Z-score)

<table>
<thead>
<tr>
<th></th>
<th>Z</th>
<th>SM</th>
<th>FO</th>
<th>CL</th>
<th>SC</th>
<th>AR</th>
<th>LE</th>
<th>OT</th>
<th>VS</th>
<th>VF</th>
<th>VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM</td>
<td></td>
<td>6.34*</td>
<td>-2.29</td>
<td>-2.36</td>
<td>2.83*</td>
<td>-1.58</td>
<td>0.11</td>
<td>2.41*</td>
<td>-0.71</td>
<td>4.02*</td>
<td>2.13*</td>
</tr>
<tr>
<td>FO</td>
<td>-5.45</td>
<td>-0.69</td>
<td>12.14*</td>
<td>-2.49</td>
<td>-0.01</td>
<td>-3.82</td>
<td>-8.86</td>
<td>-1.54</td>
<td>-3.96</td>
<td>-6.17</td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td>-0.88</td>
<td>5.42*</td>
<td>-6.22</td>
<td>-0.22</td>
<td>4.66*</td>
<td>1.45</td>
<td>-3.58</td>
<td>0.25</td>
<td>-0.78</td>
<td>-1.05</td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>0.82</td>
<td>-2.4</td>
<td>-0.18</td>
<td>-0.3</td>
<td>-1.05</td>
<td>1.64</td>
<td>2.64*</td>
<td>5.42*</td>
<td>1.96*</td>
<td>-0.63</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>1.1</td>
<td>6.48*</td>
<td>-5.57</td>
<td>-0.92</td>
<td>-3.19</td>
<td>0.02</td>
<td>-1.5</td>
<td>0.96</td>
<td>-1.3</td>
<td>2.72*</td>
<td></td>
</tr>
<tr>
<td>LE</td>
<td>0.13</td>
<td>-2.12</td>
<td>-0.13</td>
<td>2.01*</td>
<td>-0.72</td>
<td>8.53*</td>
<td>0.45</td>
<td>-0.49</td>
<td>1.13</td>
<td>-0.88</td>
<td></td>
</tr>
<tr>
<td>OT</td>
<td>1.73</td>
<td>-8.42</td>
<td>-2.35</td>
<td>1.84</td>
<td>-1.99</td>
<td>0.38</td>
<td>20.35*</td>
<td>-0.94</td>
<td>3.56*</td>
<td>2.21*</td>
<td></td>
</tr>
<tr>
<td>VS</td>
<td>3.1</td>
<td>3.42</td>
<td>3.56*</td>
<td>1.06</td>
<td>4.48*</td>
<td>2.21*</td>
<td>10.63*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VF</td>
<td>3.97*</td>
<td>-2.21</td>
<td>-1.75</td>
<td>2.31*</td>
<td>-1.32</td>
<td>-0.61</td>
<td>1.7</td>
<td>-0.44</td>
<td>3.28*</td>
<td>4.48*</td>
<td></td>
</tr>
<tr>
<td>VM</td>
<td>2.89*</td>
<td>-4.05</td>
<td>-2.39</td>
<td>1.3</td>
<td>-0.81</td>
<td>0.28</td>
<td>2.89*</td>
<td>-0.65</td>
<td>3.16*</td>
<td>10.63*</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05.

Figure 2. Overall students’ engagement pattern
In sum, behavior can be divided into two broad, independent sections based on the sequential relationships between behaviors. One type describes students’ higher engagement pattern (reciprocal relations of verbal, nonverbal, and OT behaviors), and the other describes students’ lower engagement pattern (reciprocal relations between nonverbal and nonverbal behaviors). Note that the two categories may overlap; however, we are more concerned with illustrating the differences in the degree of engagement in learning tasks rather than defining the categories.

Moreover, the students were observed murmuring after their eyes followed the arrow path (AR→VM). The students murmured words such as “Oops! Dropped it again” or “I cannot get it.” This finding may be due to our design mechanism, in which the students had to drag the garbage to the trash can; the students needed to exert more effort to control the mouse for the correct operation. This operation problem may explain some of the continuous frustration (VF→VF), reflecting two students’ comments that “the mouse is so difficult to use.”

**Gender differences**

Male students’ behavior patterns are shown in Table 4 and Figure 3. Connections between some verbal and nonverbal behaviors were observed (VF→SM, SM→VF, VM→SC, and SC→VS). Additionally, connections were found between verbal behavior and OT (VS→OT, OT→VF) and between nonverbal and nonverbal (FO→CL, CL→FO, CL→AR, AR→FO) and verbal and verbal (VS→VS, VF→VF, VM→VM, VF→VM) behaviors. However, the reciprocal connection between verbal, nonverbal, and OT behaviors occurred less frequently in male students’ behavior patterns.

On the other hand, as shown in Table 5 and Figure 4, we observed more complex connection behaviors in female students. For example, the connections between some verbal and nonverbal behavior were observed (VS→SM, VF→SM, SM→VF, VM→SM, SM→VM, VF→SC, SC→VF, SC→VS). Additionally, the connections between verbal behavior and OT were found (OT→VF, OT→VM, VM→OT). Moreover, the connections of nonverbal to nonverbal (SM→SM, SM→SC, FO→CL, CL→FO, FO→AR, AR→FO, CL→LE, SC→LE) behaviors were revealed. However, verbal to verbal behaviors were only found between VF and VM.

As indicated in Figure 5, both male and female students exhibited the same sequential behaviors while executing the game task. For example, they expressed bidirectional sequences between expressions of frustration and murmuring (VF→VM, VM→VF) and between smiling and expressions of frustration (SM→VF, VF→SM). Additionally, a continuing smile was observed (SM→SM), and students’ self-questioning behavior occurred after their scratching behavior (SC→VS).

The bidirectional sequence between focus and moving the body closer to the screen was found in both male and female students (FO→CL, CL→FO). In particular, the behavioral sequence “eyes following the narrow path to focus the element of the game” (AR→FO) was observed in both male and female students.

Different sequential behaviors between male and female students are shown in Table 4, Table 5, Figure 3, and Figure 4. Male students’ tendency to repeat verbal behaviors was observed. For example, male students tended to express self-questioning, express frustration, and murmur continually (VS→VS, VF→VF, VM→VM). However, it appeared that female students expressed more bidirectional sequences between verbal and nonverbal behaviors, as in the connections between self-questioning, expressing frustration, murmuring, and smiling (VS→SM, VF→SM, SM→VM) as well as between expressing frustration and scratching (VF→SC) (the connection was observed only between self-questioning and smiling (VF→SM) and murmuring to scratching (VM→SC)). Male students were observed moving their bodies away from the screen repeatedly (LE→LE). The leaving behavior was followed by the scratching behavior (LE→SC). Subsequently, the male students asked themselves questions after scratching their face, head, or body (SC→VF). In contrast, female students were observed moving their bodies away from the screen after scratching their face, head, or body (SC→LE).

Moreover, male students were observed moving their bodies to the screen with their eyes focusing on the arrow (CL→AR). However, this engagement pattern was rarely found in female students, who tended to focus on the elements of the game with their eyes following the arrow to execute trash classification (FO→AR). Female students also showed extreme caution while executing the game task. For example, they focused on one element and grabbed...
it on the main screen. Then, they carefully dragged it to the garbage bin while carefully watching the arrow, in case they lost the element.

**Table 4. Adjusted residuals results of male students (Z-score)**

<table>
<thead>
<tr>
<th>Z</th>
<th>SM</th>
<th>FO</th>
<th>CL</th>
<th>SC</th>
<th>AR</th>
<th>LE</th>
<th>OT</th>
<th>VS</th>
<th>VF</th>
<th>VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM</td>
<td>6.37*</td>
<td>−0.76</td>
<td>−2.3</td>
<td>−0.32</td>
<td>−1.49</td>
<td>0.48</td>
<td>1.65</td>
<td>−0.65</td>
<td>2.45*</td>
<td>0.69</td>
</tr>
<tr>
<td>FO</td>
<td>−3.77</td>
<td>−2.33</td>
<td>10.08*</td>
<td>−1.32</td>
<td>0.65</td>
<td>−2.87</td>
<td>−5.59</td>
<td>−0.84</td>
<td>−2.76</td>
<td>−4.45</td>
</tr>
<tr>
<td>CL</td>
<td>−1.36</td>
<td>2.06*</td>
<td>−3.14</td>
<td>0.04</td>
<td>5.4*</td>
<td>−0.07</td>
<td>−3.28</td>
<td>0.07</td>
<td>−0.72</td>
<td>−0.81</td>
</tr>
<tr>
<td>SC</td>
<td>−0.36</td>
<td>−0.48</td>
<td>−0.33</td>
<td>−0.12</td>
<td>−0.74</td>
<td>−0.29</td>
<td>1.57</td>
<td>4.15*</td>
<td>−0.24</td>
<td>−0.42</td>
</tr>
<tr>
<td>AR</td>
<td>1.56</td>
<td>7.68*</td>
<td>−5.38</td>
<td>−0.63</td>
<td>−4.03</td>
<td>−0.15</td>
<td>−1.5</td>
<td>0.55</td>
<td>−1.31</td>
<td>1.86</td>
</tr>
<tr>
<td>LE</td>
<td>0.52</td>
<td>−1.9</td>
<td>−0.6</td>
<td>3.78*</td>
<td>−0.92</td>
<td>9.06*</td>
<td>−0.17</td>
<td>−0.51</td>
<td>1.42</td>
<td>−0.92</td>
</tr>
<tr>
<td>OT</td>
<td>0.91</td>
<td>−5.06</td>
<td>−1.93</td>
<td>−0.45</td>
<td>−2.04</td>
<td>0.77</td>
<td>15.04*</td>
<td>−0.9</td>
<td>2.51*</td>
<td>−0.27</td>
</tr>
<tr>
<td>VS</td>
<td>0.77</td>
<td>0.28</td>
<td>−1.61</td>
<td>0.23</td>
<td>−1.44</td>
<td>−0.57</td>
<td>3.32*</td>
<td>4.1*</td>
<td>−0.47</td>
<td>−0.82</td>
</tr>
<tr>
<td>VF</td>
<td>2.51*</td>
<td>−1.6</td>
<td>−0.74</td>
<td>−0.21</td>
<td>−1.34</td>
<td>−0.53</td>
<td>1.39</td>
<td>−0.42</td>
<td>4.26*</td>
<td>2.03*</td>
</tr>
<tr>
<td>VM</td>
<td>0.74</td>
<td>−2.37</td>
<td>−1.91</td>
<td>2.52*</td>
<td>−1.31</td>
<td>0.23</td>
<td>0.45</td>
<td>−0.73</td>
<td>2.03*</td>
<td>9.43*</td>
</tr>
</tbody>
</table>

*p < 0.05.

**Table 5. Adjusted residuals results of female students (Z-score)**

<table>
<thead>
<tr>
<th>Z</th>
<th>SM</th>
<th>FO</th>
<th>CL</th>
<th>SC</th>
<th>AR</th>
<th>LE</th>
<th>OT</th>
<th>VS</th>
<th>VF</th>
<th>VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM</td>
<td>2.47*</td>
<td>−2.53</td>
<td>−0.98</td>
<td>3.62*</td>
<td>−0.72</td>
<td>−0.54</td>
<td>1.77</td>
<td>−0.31</td>
<td>3.35*</td>
<td>3.35*</td>
</tr>
<tr>
<td>FO</td>
<td>−4.14</td>
<td>−1.25</td>
<td>7.97*</td>
<td>−2.57</td>
<td>2.98*</td>
<td>−2.14</td>
<td>−6.98</td>
<td>−0.84</td>
<td>−2.82</td>
<td>−3.49</td>
</tr>
<tr>
<td>CL</td>
<td>0.18</td>
<td>6.48*</td>
<td>−5.99</td>
<td>−0.2</td>
<td>−1.96</td>
<td>2.59*</td>
<td>−1.73</td>
<td>0.2</td>
<td>−0.39</td>
<td>−1.14</td>
</tr>
<tr>
<td>SC</td>
<td>1.19</td>
<td>−3</td>
<td>0.09</td>
<td>−0.32</td>
<td>−0.45</td>
<td>2.71*</td>
<td>2.31*</td>
<td>5.06*</td>
<td>2.71*</td>
<td>−0.34</td>
</tr>
<tr>
<td>AR</td>
<td>−0.75</td>
<td>3.53*</td>
<td>−2.56</td>
<td>−0.4</td>
<td>−0.57</td>
<td>−0.43</td>
<td>−0.89</td>
<td>−0.25</td>
<td>−0.43</td>
<td>−0.43</td>
</tr>
<tr>
<td>LE</td>
<td>−0.53</td>
<td>−0.38</td>
<td>0.55</td>
<td>−0.28</td>
<td>−0.4</td>
<td>−0.3</td>
<td>1.06</td>
<td>−0.17</td>
<td>−0.3</td>
<td>−0.3</td>
</tr>
<tr>
<td>OT</td>
<td>1.62</td>
<td>−6.89</td>
<td>−1.44</td>
<td>2.74*</td>
<td>−0.89</td>
<td>−0.67</td>
<td>13.58*</td>
<td>−0.38</td>
<td>2.51*</td>
<td>5.69*</td>
</tr>
<tr>
<td>VS</td>
<td>5.23*</td>
<td>−1.28</td>
<td>−1.27</td>
<td>−0.2</td>
<td>−0.28</td>
<td>−0.21</td>
<td>1.93</td>
<td>−0.12</td>
<td>−0.21</td>
<td>−0.21</td>
</tr>
<tr>
<td>VF</td>
<td>3.19*</td>
<td>−1.42</td>
<td>−1.91</td>
<td>3.1*</td>
<td>−0.43</td>
<td>−0.32</td>
<td>0.92</td>
<td>−0.18</td>
<td>−0.32</td>
<td>6.1*</td>
</tr>
<tr>
<td>VM</td>
<td>5.06*</td>
<td>−2.77</td>
<td>−1.91</td>
<td>−0.3</td>
<td>−0.43</td>
<td>−0.32</td>
<td>5.69*</td>
<td>−0.18</td>
<td>2.89*</td>
<td>−0.32</td>
</tr>
</tbody>
</table>

*p < 0.05.

**Figure 3. Sequential engagement patterns of male students’ behaviors in game-based learning environments**
Regarding gender issue in game-based learning, a number of commentators believe that gender difference may be diminishing or even disappearing (Ke & Grabowski, 2012; Papastergiou, 2009). However, when considering the learning process, a difference in the learning pattern was evident. For example, male students were observed demonstrating verbal behaviors when they encountered problems, whereas female students exhibited both verbal and nonverbal behaviors when they were confused. Male students demonstrated more engaged behaviors with continuous self-conversations, whereas the frequency of self-conversation by female students was obviously lower than those of male students. The result is in line with previous studies suggesting that males were more involved in the game task and had greater engagement in exploratory behavior than females (Papastergiou, 2009). Moreover, males were found to be more competitive than females and preferred to test themselves in games of skill (Inal & Cagiltay, 2007). In contrast, female students may begin thinking with facial expressions or body language, and these nonverbal behaviors may help to release their tension. This finding may indicate that female students are less confident about using technology to execute learning tasks and need to give themselves space to think. Inherent differences between male and female students’ learning was observed in this study.
Conclusion and suggestions

By connecting obvious nonverbal and verbal behaviors, this study visualized the learning process and provided evidence that the game can consistently increase students’ engagement in the game-based learning environment, which can provide insights as a good start for qualitative study in game-based learning. Thus, the behavioral coding schema provided by the authors and the innovative method of sequential analysis may help researchers obtain a certain level of understanding of students’ engagement patterns and provide an alternative reference for researchers. The results of this study encourage researchers to use this innovative method.

The sequential patterns represented qualitative similarities and differences of engagement patterns by gender. For example, both male and female students exhibited the same sequential behaviors, such as expressing frustration, murmuring continually, or smiling. Differences in engagement patterns were also observed. For example, male students often demonstrated more engaged behaviors with continuous self-conversation, whereas female students did so noticeably less frequently. They tended to present both verbal and nonverbal behaviors when they were confused.

Observing how students respond to conflicting questions during the gameplay process provides an important modeling opportunity for educators to supply appropriate and effective facilitation for students. This may help educators and researchers to develop better game mechanisms to help students engage in meaningful learning. Moreover, these behaviors seem to be subconscious. We believe that follow-up interviews may help to explain the reasons for these behaviors. Interviews may include student-initiated questions and discussions of how they explored difficult concepts during the gameplay process as well as why students decided to terminate their engagement and failed to maintain attention to the game process.

This study proposes suggestions and notes limitations for future researchers who are implementing sequential analysis in game-based learning environments.

• In this study, students only had approximately 15–20 minutes to execute the game. The time participants spent conducting the game task may influence the grounding categories of verbal and nonverbal behaviors. Future studies should give a habituation time to avoid novelty effects. Thus, we suggest that future research should give students a longer time to play to allow researchers to collect data showing long-term changes in behavioral data and to explore the possible relationships between the behaviors. Moreover, the number of participants in this study was rather small. Future studies should increase the number of participants to improve the results.

• Finally, some elementary school students may not be able to control the mouse and may find it difficult to press and hold down the mouse button. Because poor usability is a barrier to engagement (O’Brien & Toms, 2008), this study suggests that future research could use keyboards instead of mice, especially for younger students. Moreover, the elementary school students’ learning behaviors may be less constrained by the environment compared with elder students’. Thus, future process studies should recognize the limitations of inference from current field practice. We suggest further studies to explore possible engagement behaviors in elder students.

Acknowledgments

This research was supported by the projects from the Ministry of Science and Technology, Republic of China, under contract number MOST-102-2511-S-011-001-MY3 and MOST-100-2628-S-011-001-MY4.

References


Understanding the Self-Directed Online Learning Preferences, Goals, Achievements, and Challenges of MIT OpenCourseWare Subscribers

Curtis J. Bonk1*, Mimi Miyoung Lee2, Xiaojing Kou3, Shuya Xu1 and Feng-Ru Sheu4

1Department of Instructional Technology and Educational Psychology, Indiana University, 201 N. Rose Avenue, Bloomington, IN 47405-1006, USA // 2Department of Curriculum and Instruction, University of Houston, 429 Farish Hall, College of Education, Houston, TX 77204, USA // 3Center for Language Technology, Indiana University, 1020 E. Kirkwood Ave., Bloomington, IN 47405-7104, USA // 4University Libraries, Kent State University, 1125 Risman Drive, Kent, OH 44242-0001, USA // cjbonk@indiana.edu // mlee7@uh.edu // xkou@indiana.edu // xushuy@indiana.edu // fsheu@kent.edu

*Corresponding author

(Submitted June 13, 2014; Revised September 19, 2014; Accepted October 23, 2014)

ABSTRACT

This research targeted the learning preferences, goals and motivations, achievements, and possibilities for life change of self-directed online learners who subscribed to the monthly OpenCourseWare (OCW) e-newsletter from MIT. Data collection included a 25-item survey of 1,429 newsletter subscribers; 613 of whom also completed an additional 15 open-ended survey items. The 25 close-ended survey findings indicated that respondents used a wide range of devices and places to learn for their self-directed learning needs. Key motivational factors included curiosity, interest, and internal need for self-improvement. Factors leading to success or personal change included freedom to learn, resource abundance, choice, control, and fun. In terms of achievements, respondents were learning both specific skills as well as more general skills that help them advance in their careers. Science, math, and foreign language skills were the most desired by the survey respondents. The key obstacles or challenges faced were time, lack of high quality open resources, and membership or technology fees. Several brief stories of life change across different age ranges are documented. Among the chief implications is that learning something new to enhance one’s life or to help others is often more important than course transcript credit or a certificate of completion.

Keywords

Open education, OpenCourseWare, Self-directed learning, Massive open online courses (MOOCs), Informal learning, Motivation

Introduction

Twenty-first century learners now have a seemingly endless array of open educational resources (OER), open courseware (OCW), and massive open online courses (MOOCs) available for their self-directed learning pursuits. For example, Wendy Ermold, a researcher and field technician for the University of Washington Polar Science Center conducts research in remote northern regions of the world (Bonk, 2009). Wendy informed us that when out on the icebreakers or remote islands, she listens to lectures and reviews various OER she has found. The content she uses often comes from the MIT OCW project as well as from Stanford University, Seattle Pacific University, and Missouri State University. As such free and open educational resources expand, learning becomes increasingly personalized.

Hundreds of millions of people like Wendy are now learning using some online tool, resource, or activity each day. As such, the web offers new hope for learning a hobby, obtaining a certificate or degree, or pursuing some other personal lifelong learning dream. Unfortunately, there are few research projects documenting the results of the use of open educational content (Iiyoshi & Kumar, 2008). There is a need to find out what motivates people to learn informally from open educational content and what skills they are attempting to learn. It is also vital to document the challenges and obstacles such learners face. At the same time, case studies of individuals whose lives have been altered or significantly changed from such casual informal learning online as well as more extended informal learning pursuits can serve as inspirational stories to others.

This paper documents the informal learning experiences and preferences of those using open educational contents as well as their motivations, achievements, and obstacles. It also captures a few samples of life changes or “empowerment moments” across different age ranges from open educational opportunities. By cataloguing some of
the ways in which informal and nontraditional web-based learning have impacted people across ages, gender, ethnicities, and cultures, we hope to encourage others to continue to learn across the lifespan. This study is among the first steps in this process.

**The OpenCourseWare (OCW) initiative**

On April 4, 2001 (i.e., 4-4-1), Charles Vest, then president of MIT, made an historic announcement with the explicit goal of having most of his prestigious university’s courses on freely available on the web within a decade (MIT News, 2001). A little over a month later, on May 25, 2001, CCN reporter, Richard Stenger (2001), reminded us that it was the fortieth anniversary of the speech by former U.S. President John F. Kennedy wherein he offered a similar bold challenge to get a person to the moon and back by the end of the 1960s. Like Kennedy’s dream of space travel before him, some educators thought Vest’s speech was a rather bold and unattainable proclamation. Nevertheless, by the early part of 2009, MIT had its entire curriculum of over 1,800 courses online. In fact, MIT beat its original target by more than 3 years (MIT, 2007). Today, all of MIT’s courses are not only available for self-directed learners around the globe to explore, download, use, and share, but for high school students in advanced placement courses in sciences, mathematics, engineering, and the humanities and social sciences as well as for those taking massive open online courses (MOOCs). Importantly, these courses are continually updated, enhanced, and expanded upon.

Vest had thought that the Council on Educational Technology that he had tasked with investigating online learning and opportunities outside classroom walls would come up with new revenue models for MIT content. He did not envision that MIT would be giving away his contents on the web. At the same time, he thought that the OCW project would be highly innovative and help advance education by widening access to it and inspiring other institutions of higher learning to also participate. As Vest (MIT News, 2001) noted, “This is about something bigger than MIT. I hope other universities will see us as educational leaders in this arena, and we very much hope that OpenCourseWare will draw other universities to do the same. We would be delighted if -- over time -- we have a world wide web of knowledge that raises the quality of learning -- and ultimately, the quality of life -- around the globe.”

Vest viewed the OCW initiative as one that embraced the openness of education and outreach to underserved and disadvantaged populations as well as professionals in the workplace, college professors seeking innovative teaching examples, and retirees wanting to learn or experience new hobbies. Learners could draw upon these materials for self-study and self-improvement. At the same time, instructors soon could share contents through a consortium OCW types of projects on other campuses around the world (Carson, 2009; Caswell, Henson, Jensen, & Wiley, 2007). With more than 2.2 million visitors to the OCW website each month and translation sites receiving hundreds of thousands of additional visits (S. E. Carson, personal communication, January 14, 2014), there is no doubt that Vest was correct in assuming that there was a large population interested in such content.

**The rise of online self-directed learning**

With the advancement of various learning technologies combined with the opening up of educational resources, issues related to self-directed learning have gained much interest in various fields. Self-directed learning is defined as “a learner’s autonomous ability to manage his or her own learning process, by perceiving oneself as the source of one’s own actions and decisions as a responsibility towards one’s own lifelong learning” (Sze-Yeng & Hussian 2010, p. 1913). Stated another way, self-directed learners take initiative related to their own learning with or without an instructor present. Prominent adult educator, Brookfield (2013) places emphasis in learners deciding on what to learn, when to learn it, how much to learn, and whether something has been learned well enough. From his perspective, the truly self-directed learner is empowered, not controlled by external decisions to acquire predetermined skills or negotiate through some heavily structured curricula. Learning decisions rest with the learner.

As indicated, online learning and free and open contents have also transformed many aspects of life for adult learners. For instance, through OCW, OER, and now MOOCs, those stuck behind prison walls, injured and in a hospital bed, or unemployed and unable to pay for college tuition can learn to be more productive members of society. Some people might be in transition from one career to another and find OER and OCW can arouse new
interests and confidence (Iiyoshi & Kumar, 2008). Still others might be enrolling in online courses or MOOCs while in war zones in Iraq or Afghanistan (Kenning, 2012).

As society moves from an age of information scarcity to one of abundance, opportunities for learners to informally learn and self-direct important aspects of their learning lives have emerged. Decades of research from Edward Deci and Richard Ryan on self-determined learning (Deci & Ryan, 2012), for instance, has highlighted the need for learners to engage in tasks that they find personally meaningful, interesting, and enjoyable across the lifespan. From their viewpoint, learners who are self-determined have a high sense of autonomy, competence, and relatedness to others. Through personal goal-setting, they self-monitor and evaluate their learning progress (Reeve, 1996). In effect, they take responsibility for their own learning.

Such views parallel one’s from humanist Carl Rogers (1983) who argued that humans learn best in environments that involve much choice, respect learner interests, and are highly collaborative. He believed that those who felt a sense of freedom and openness to new experiences would become more expressive and creative. From this perspective, learner participation is emphasized over learning consumption of lectures and book materials.

With the growth of open education contents and learning portals on nearly any topic of importance, members of the twenty-first century now have a plethora of choices and opportunities about what they will learn. Mol and van Dam (2013) note that the future of work depends on people being able to acquire the skills that they need when they need them. They further contend that open educational opportunities such as MOOCs offer a powerful way to reskill the workforce and build expertise from the cloud with an on-demand training program. Given the expanding growth of online information and educational resources, there is a mounting need for research on self-directed learning (Hyland & Kranzow, 2011).

**Research on self-directed learning**

There has been increasing research attention on self-directed learning during the past decade in both formal and informal learning environments. For instance, based on a course designed using a self-directed learning approach, a study by Sze-Yeng and Hussain (2010) attempted to foster self-directed learning of instructional technology master’s students. To accomplish this goal, they employed a blended learning environment using Moodle, Google Docs, and Wikispaces based on socio-constructivist principles. Their design decisions involved students becoming aware of their own existing knowledge and further expanding on it through interaction with peers and teachers.

Self-directed online learning is essential today in most professional fields, especially in healthcare (Li, Tancredi, Co, & West, 2010). For instance, El-Glinay and Abusaad (2013) investigated the self-directed learning readiness and learning styles among undergraduate nursing students. They noted that the reason for the interest in self-directed learning in nursing education during the past few decades is due to the increasing complexity and myriad changes within the nursing profession. Importantly, they found no relationship between self-directed learning readiness (SDLR) and learning styles, nor was it related to age, gender, marital status, or urban or rural residence. Such findings, in effect, imply that there is much opportunity to learn through self-direction across the lifespan.

In a national survey of pediatric medical residents and program directors, Li et al. (2010) attempted to identify the factors that contributed to successful self-directed learning. The results revealed that the learner-level factors such as using a system to track one’s progress in achieving learning goals, a propensity toward lifelong learning, and higher levels of confidence in self-directed learning, as compared to the program-level factors, were more vital to the success of these self-directed learners.

In terms of research on informal learning, attempts have been made to evaluate the effectiveness of informal learning in the workplace (Jeon & Kim, 2012) as well as the value of the interactions that are happening in the informal learning environments such as YouTube or Facebook (Tan, 2013). Over the past few years, informal learning in the workplace has been gaining increasing attention and interest due to the need for prompt updates on skills and information in most jobs. New forms of learning and literacy are forming at the intersection of academic needs and informal learning sources including unique spaces like the Khan Academy and other YouTube channels (Goodfellow, 2011; Meyers, 2014). Such resources challenge traditional forms of learning and interaction (Tan, 2013).
Purpose and method

The purpose of this study is to explore the self-directed and informal learning experiences of subscribers to the monthly MIT OpenCourseWare (OCW) online newsletter. In particular, the research targeted the (1) learning preferences, (2) goals and motivations, (3) achievements, (4) obstacles and challenges, and (5) possibilities for life change of self-directed online learners. With greater understanding of the goals and successes as well as the obstacles that self-directed online learners face, educators and instructional designers can design and develop more relevant and potentially powerful online learning contents and instructional scaffolds.

Prior to the study, a list of over 300 informal and extreme learning websites related to language learning, social change, global education, virtual education, adventure learning, shared online video, and open education in general had been generated and then evaluated by a team of researchers using an eight-part coding scheme (Jung, Kim, Wang, & Bonk, 2011; Kim, Jung, Altuwaijri, Wang, & Bonk, 2014). During the evaluation process, the researchers noted a wide array and diversity of informal learning experiences, skills or competencies emphasized, delivery mechanisms and technologies utilized, motivational techniques employed, and potential barriers or obstacles to their use. One particularly interesting website that the team evaluated was the OCW project from MIT. Consequently, MIT OCW officials were contracted to participate in this study. Using insights from the website evaluation process, a 40-item survey (see http://trainingshare.com/pdfs/MIT-survey.pdf) was designed that took MIT OCW participants around 15 to 20 minutes to complete (to simulate taking the survey, see http://www.surveyshare.com/s/AYA4CTD).

The close-ended portion of the survey inquired into many aspects of informal learning: the goals one wished to accomplish through informal learning pursuits and activities, the reasons for exploring web resources informally, the factors leading to success, what they would like to learn informally online, and the typical barriers or obstacles faced when learning informally on the web. We also asked a question about what they would like to achieve from their informal learning online.

In addition to the initial 25 close-ended questions, respondents had the option to complete 15 open-ended questions. The open-ended questions included those related to goals and aspirations as well as challenges and obstacles and possible life changes using OER, OCW, and MOOCs. Participants were also asked about their most interesting and successful informal learning experiences and what they accomplished.

After qualitative researchers on our team analyzed the data multiple times, a decision was made to examine the 15 answers for each survey participant; in effect, treating these answers as one short interview per respondent. As a result, respondent goals and motivations for utilizing OCW and other forms of open education might be apparent in the question that specifically asked about it as well as in other question responses. Two rounds of coding produced the necessary coding schemes. The qualitative data presented here are intended to make evident how lives are impacted from open education across the lifespan.

Participants

The research data was collected in August 2012 through a web-based survey. The sample was derived from subscribers to the e-newsletter related to the popular MIT OCW initiative. At the time, the newsletter subscription list had more than 156,000 active subscribers, of which, some 41 percent were described as self-learners, 40 percent students, 15 percent educators, and 3 percent parents and other users (MIT OpenCourseWare, 2012). About 26,700 people opened the email and 4,000 people clicked through to the survey. Some 1,429 people completed the survey, including 613 people who completed one or all of the optional open-ended survey items.

In terms of age, about half of the 1,429 survey respondents were age 40 or younger. Interestingly, 64 respondents were over age 70 which equated to roughly 5 percent of the MIT OCW sample pool. In addition, most in the MIT sample were males (76 percent). While nearly half of the respondents were from North America (618 people; 44 percent), a significant number came from Asia (331 people; 23 percent), Europe (202 people; 14 percent), and South America (133 people; almost 10 percent). Among the top countries represented in the MIT OCW subscriber list were the United States, India, China, Brazil, Nigeria, Pakistan, Iran, Canada, the UK, Taiwan, Indonesia, Mexico, and Egypt.
Findings from close-ended survey questions

*Learning preferences results.* The survey results indicated that the respondents typically used a laptop (83 percent) or desktop (72 percent) to access informal learning resources, though some used a smartphone (38 percent), tablet computer (29 percent), or e-book reader (29 percent). Respondents were also learning from devices such as iPods, car CD players, and the Internet on their televisions.

We also inquired where these self-directed learners were learning informally. Among the popular places for accessing informal learning resources and materials were at home (91 percent), work (44 percent), school or university (40 percent), libraries (36 percent), anywhere with a mobile device (36 percent), and cafes and bookstores (26 percent). However, as shown in Figure 1, web access from airports, buses, subways, and trains were also among the common places respondents went online to learn something.

![Figure 1. Where self-directed online learners engage in informal learning](image)

We asked two questions about specific tools and web resources self-directed online learners used when seeking new skills, information, or answers to their questions. First, we asked the respondents to list the three best websites—other than search engines like Bing or Google—that they used when they had a fairly simple task or question. Of the 1,429 survey respondents, 1,237 provided at least one answer and 785 of them listed three as requested. The overwhelming majority (715 people) listed Wikipedia as their primary source for their basic information needs. The second most popular website was the MIT OCW project which all these people had subscribed to (256 people). Third in line was YouTube (170 people). Next came the Khan Academy (73 people), How Stuff Works (54 people), Wolfram Alpha (48 people), eHow (46 people), and Ask.com (40 people). In addition, between 25 and 36 people listed About.com, Stack Overflow, NPTEL (National Program on Technology Enhanced Learning) India, Coursera, the New York Times, WebMD, and Yahoo! Answers as resources that they used to address such basic knowledge question needs.

The following question asked the respondents to list the three best educational or information-rich websites that they “might recommend to others that can significantly influence or change their lives in a positive way.” Interestingly, the main life changing resource among these subscribers to the MIT OCW newsletter was MIT OCW itself (649 people). Third in line was YouTube (170 people). Next came the Khan Academy (73 people), How Stuff Works (54 people), Wolfram Alpha (48 people), eHow (46 people), and Ask.com (40 people). In addition, between 25 and 36 people listed About.com, Stack Overflow, NPTEL (National Program on Technology Enhanced Learning) India, Coursera, the New York Times, WebMD, and Yahoo! Answers as resources that they used to address such basic knowledge question needs.

The following question asked the respondents to list the three best educational or information-rich websites that they “might recommend to others that can significantly influence or change their lives in a positive way.” Interestingly, the main life changing resource among these subscribers to the MIT OCW newsletter was MIT OCW itself (649 people). In effect, there apparently was plenty of potential for life change embedded in the very resources that they were relying on. The other three websites that more than 100 survey respondents listed were Wikipedia (289 people), the Khan Academy (145 people), and Coursera (144 people). Other highly popular resources included TED (84 people), YouTube (67 people), and Udacity (55 people). Resources with at least 25 votes included Stanford (presumably, the Stanford Venture Lab which offered a series of MOOCs), iTunes University, NPTEL India, edX, the BBC, and How Stuff Works.
**Goals and motivations results.** While grasping the learning preferences related to the external factors impacting self-directed learning such as devices, locations, and tools is illuminating, it is equally or more important to know the learner goals, needs, and overall motivations when engaging in informal and self-directed online learning. As shown in Figure 2, intrinsic motivation trumped extrinsic motivation for these self-directed learners. More specifically, curiosity, seeking information, self-improvement, and wanting to learn something were the key reasons to informally explore the web to learn. Nearly 70 percent, in fact, had personal goals for self-improvement. A similar percentage were simply satisfying their curiosity. More impressively, nearly 80 percent entered the web to find out about a particular topic (see Figure 3). Also of interest was that more than half of the respondents were learning online for professional development reasons. Similarly, 54 percent used the web because they wanted more information and 57 percent wanted to learn something new. Some wanted personal control over their learning (46 percent) or personally felt that they needed new skills (46 percent). Fewer respondents were exploring the web to help with their hobbies (35 percent) or because they wanted to make a contribution to society (27 percent).

![Figure 2. Main reasons to informally explore the web to learn](image)

The researchers specifically asked about key factors that typically led to their online learning successes when engaging in informal online experiences (see Figure 3). As predicted by humanistic psychologists like Rogers (1983), freedom to learn was rated most important to their success (72 percent). The next most important factors were a sense of resource abundance (47 percent), choice (44 percent), control over the activity or resource (41 percent), sense of fun (40 percent), and producing or creating something new (37 percent). Factors like system support, feedback, novelty, sharing, collaboration, and sense of adventure were also valued in different degrees.

Participants were asked what they would like to achieve from their informal learning endeavors (see Figure 4). While nearly 85 percent sought out informal online learning for a new skill or competency, 57 percent were there to engage in a learning experience that would better their life. Some wanted to simply fix something at home (43 percent), whereas others had more grandiose goals of helping society (47 percent), learning something to help others (53...
percent), or acquiring information about cultures or communities in the world (41 percent). About 46 percent were actually hoping to obtain course credit, while 38 percent wanted a course or module but did not care if it led to a degree.

**Figure 3.** Factors leading to success or personal change what learning informally online

**Figure 4.** What would you like to achieve from informal learning online
In terms of specific skills, it is not too surprising that people seeking access to course resources from MIT would be interested in learning mathematics and science related content. In fact, more than 60 percent of the respondents were seeking mathematics content and over 75 percent were interested in science (see Figure 5). The only other category above 50 percent was foreign language at 57 percent. Given the emphasis on math and science skills as would be expected from users of MIT OCW, it was interestingly that more than 4 in 10 people were seeking global information (45 percent) and cultural information (42 percent). In addition, nearly 4 in 10 were accessing these resources for historical content (39 percent) or health-related information (37 percent). And about one-third of them were seeking new vocabulary skills as well as environmental information. Art and music information or resources were only sought out by 1 in 4 people. Even fewer were interested in acquiring outdoor skills (15 percent) or athletic skills (10 percent). Clearly, there were many different types of learning pursuits from these open educational resources provided by MIT and others.

![Figure 5. Specific skills respondents wanted to learn informally online](image)

**Learning achievements results.** Not surprisingly, there were a range of actual accomplishments from informal learning pursuits (see Figure 6). Nearly everyone admitted to learning something new (88 percent). The next most selected response was feeling better about oneself as a learner after their open education experiences. More specifically, over 6 in 10 respondents indicated that their identity as a learner was enhanced (61 percent). Approximately half of the respondents indicated that their personal freedom, in general, improved. Over 40 percent felt better about themselves as human beings. Importantly, in the process, more than 37 percent changed their beliefs about learning as a result of informal learning pursuits. As might be expected, a fairly large percentage became interested in a new occupation or career (36 percent). More impactfully, perhaps, one in five claimed to have actually found a new job or position as a result of their informal learning experiences. A similar percentage received a certificate of some kind. Clearly, informal online learning had a powerful effect on the MIT OCW participants.

Less important, it seems, was keeping up with one’s friends or making new friends. And while not huge, it is important to note that slightly more than 10 percent indicated that they moved up at work. Around the same percent of respondents noted that their self-directed learning pursuits helped them to score higher on standardized tests or exams.
Figure 6. Achievements from informal learning pursuits

Figure 7. Obstacles and challenges faced when learning informally online
Obstacles and challenges results. As displayed in Figure 7, while many respondents had numerous self-directed learning successes, they also encountered significant obstacles and challenges when learning informally online. For instance, the most significant issue for exactly half of the respondents was a lack of time to use the resource. Such time constraints are often noted by those enrolled in MOOCs and other time intensive online courses. Other issues might include the lack of support within one’s work environment for informal learning (17 percent), difficulty in using the site or service (23 percent), the lack of high quality open resources in a particular area (32 percent), and membership or technology fees (45 percent). In addition, 17 percent indicated a lack of access to the site or service, 14 percent had outdated or inappropriate technology, and 13 percent experienced firewall barriers. Other problems the respondents noted were a lack of personal motivation (14 percent) and lack of excitement to use the resource (9 percent).

Possibility for life change results. Among the most important findings from the survey was that the respondents experienced some type of life change from their informal learning pursuits. In fact, more than 75 percent of the respondents felt a sense of life change. Clearly, learning informally on the web has a major societal as well as personal impact. Another interesting finding was that those helping the respondents to learn informally online were friends, peers, and colleagues (43 percent), people they never met (31 percent), experts (26 percent), and teachers or trainers (23 percent). As open educational contents and courses expand online, informal learners will increasingly rely on support from experts and those who they may never meet.

Qualitative findings related to life change across the lifespan

The issues and topics detailed above related to the (1) learning preferences, (2) goals and motivations, (3) achievements, (4) obstacles and challenges, and (5) possibilities for life change of self-directed online learners were further explored in the open-ended survey responses. In this manuscript, we only report those related to the final area of life change.

One way to understand the impact of OCW is to explore the qualitative data according to different age groups. There were numerous stories of life change embedded in the open ended data. Seven different ones are briefly detailed below based on age from those in their teenage years to people now in retirement who are starting new careers. These examples were selected to illustrate the varied reasons in which people access and learn valuable information from open educational contents. These are just seven stories from more than 600 that were collected in the open-ended survey items.

Case #1. 18-20 year old male from the Middle East.

Result. OCW inspires young people and can change entire educational systems.

“When i was 14 years old i found MIT OCW during my search in..(physics) by Prof Walter Lewin looked really interesting and i became interested in physics. To be honest OCW changed my way of living and i found how beautiful physics is…informal learning is interesting because you can have access to some of the best courses provided by the best universities in the world…MIT OCW or Stanford open courses have also changed the educational system in some poor countries and have taught the teachers and professors in those countries how to teach a subject in a modern way.”

As the quote above indicates, OCW developed for college students at MIT can be used by middle and high school students who normally would not have access to that information. Such materials can inspire young people into a career that they had not previously contemplated. They can also provide a sense of curiosity and wonder about a concept, theory, or entire discipline. Such individuals might later become advocates for the field who encourage their peers to learn about the content area. At the same time, the online videos and other resources from inspirational and engaging instructors can offer inexperienced instructors a means to learn new skills and teaching techniques. As such, informal learning can impact highly formal teaching and learning.
Case #2. 21-30 year old unemployed female from North America.

Result. OCW helps people become self-taught about social media and start new careers.

“At first my purpose was to fulfill boredom …After graduating with a MS, I was faced with unemployment. I took the opportunity to read blogs, watch Youtube videos, and more to learn about blogging and social media. Since then I have become well versed in social media and other business topics and started a business… I decided not to pursue a PHD because I am learning a more rapid pace. Instead of spending 5 years in school, I can be flexible and work on what I am learning.”

Many young people today remain unemployed after obtaining a college degree and many more are uncertain about their future. As the quote above reveals, OER and OCW offer a chance for such individuals to gain new knowledge and skills as needed without having to head back to college for a lengthy commitment such as pursuing a doctorate. Just-in-time learning available through blog posts, podcast shows, online news, open access research articles, and YouTube and other shared online video resources allows self-directed learners to obtain needed information at the time and place most appropriate to their needs. Equally important, this abundance of open educational resources can accelerate the learning process. When that happens, much money can be saved on tuition. Many years previously spent in school can now be reallocated to making a contribution to society. And if success is attained, one’s personal identity and self-esteem are enhanced.

Case #3. 31-40 year old male from Asia

Result. Self-Directed learning enhances jobs skills and new perspectives.

“I learnt scheme from MIT OCW. Which helped in learning elisp/lisp. The programming techniques increased my software design knowledge. Algebra - mit ocw course was very useful, it helped me to refresh the basics of Digital signal processing. - I have learnt a lot of "applying my knowledge" than just learning the "theory" - Various online classes allow for multiple perspectives of the same topic thus showing us how the same thing can be applied in different fields.”

Revealed in the quote above is an obvious use of MIT OCW. People search for content online that can elevate their status in a present job situation. Skills inadequately learned or long forgotten can be relearned. In addition, cutting-edge content such as the latest in computer programming can be acquired as needed and immediately applied on the job. As such, the self-directed learner can test and continue to refine the newly formed skills. This individual also points out that with multiple courses available online on the same topic, he can integrate multiple perspectives or ways that instructors teach a given topic.

Case #4. 41-50 year old male from the Middle East.

Result. Better prepared to reenter university life.

“Most interesting experience of my own was my use of MIT OCW to refresh on Calculus. I purchased the textbook and followed one of several calculus options on the site. This was quite successful in re-introducing Calculus, as a prep to re-entering college 28 years after graduation - this time to study for a Master’s.”

The middle aged individual from the Middle East who is quoted above found his refresher course on Calculus to be an ideal course to prepare him for his upcoming enrollment in a master’s program. Interestingly, he points to several options that MIT OCW offered to learn Calculus; he was not restricted to one set path or set of course materials. Instead, he could pick and choose from several resources or courses. As indicated earlier, having such choices is highly motivational and can foster a greater sense of personalization of learning.

Case #5. 51-60 year old female from North America.

Result. Open educational resources and open access to well-known people builds expertise and possibilities for a learning apprenticeship.
“I'm a virtual reference librarian…My friend and I have a blog about music cognition, which is a little crazy because we don't know anything about it. Nevertheless, we blog about current articles, and one experience was especially meaningful, where I tweeted a question to Daniel Levitin, who answered. He's the guy who wrote The World in Six Songs, among other things. The fact that you can communicate with an expert in the field who will take the time to give a thoughtful tweeted (and yes, 140 characters can be thoughtful)--well, that's amazing…I think it was the single most exciting learning moment I've experienced.”

Several things are interesting about this quote. First of all, this individual decided to create a new blog in an area in which she had limited skills and knowledge. Such a bold move might indicate that learning technologies like a blog can be the virtual napkin from which to explore one’s emerging ideas and interests. What is even more interesting is that she received guidance and support from an established musical performer. In effect, in a highly self-directed world, our teachers might be people we never have met; including some we never would have imagined would respond to our queries. Feedback and coaching on our ideas can come from anywhere. This individual definitely was delighted by this unexpected support. As such events occur, learner intrinsic motivation is enhanced.

**Case #6.** 65 year old male from North America

**Result.** A retiree develops a new hobby through OER and OCW and that hobby requires still more skill.  
“I retired from education I became a lifestyle entrepreneur. I was trying to learn software to develop a business website for our sailing business. I discovered…Lynda.com. I was learning web development, video editing and photoshop. I realized that I could learn this software without going to a formal class and that was an "eureka" moment. My wife and I travelled and started a yacht delivery and training business that has taken us all over the world. We learned technology and social media as we travelled and ended up with many fantastic experienced in some very exotic places. We taught ourselves web development and have been figuring out online learning as we went. Now at 65 years of age we are going back to university to take a Computer Science Degree and start a new tech business relating to interactive educational media and games. All very exciting :-("

As with the second case, online courses and OER can equip a self-directed learner with timely skills to be a successful entrepreneur. In this case, however, it is not someone starting a new career, but, instead, a husband and wife team who are retired. Their learning of web design and development as well as related skills has helped them develop a sailing business. Based on their extensive success, they have decided to enroll in a university for a degree in computer science. In effect, their informal learning has been so fulfilling that they are enrolling in formal education. Importantly, they are able to balance their various online knowledge pursuits and their building of their new business. And they are doing so in their mid-60s. With OER and OCW, learning never ends; instead, a series of life paths can unfold across the lifespan.

**Case #7.** 70+ year old retired female from North America.

**Result.** New web design skills help a retiree showcase her talents.  
“I'm an avid photographer and have developed skills for developing my own website for display of my photography and books I have written that include my photographs…I've had multiple careers, from science, to public administration, and information technology. The development of my art is a new and exciting experience.”

This final quote illustrates the use of OER and OCW across the lifespan. Those who retire or who pursue new hobbies need multiple access points to learning. Free and open online courses or even just a few modules can provide the skills needed to run a business (e.g., tax accounting) or acquiring skills needed to market or demonstrate one’s talents as evident in this case. What is particularly interesting is that this 70+ year old former scientist and IT administrator is learning how to design a homepage on the web from course contents found on the web. Her art is a new career. With open educational materials, she can continue to develop those skills and explore new income producing and life satisfaction options.
These seven individuals provide an important glimpse into the potential for life change from open educational resources. They each represent a different age group or stage in life. Importantly, all seven experienced something significant as a result of learning online. Clearly, MOOCs, OCW, and OER can help in building, enhancing, or changing a career. Additional research is needed to ferret out the motivational components and support structures that can be embedded in open content to help larger numbers of people take advantage of open education.

**Recapping the results and conclusions**

This study explored five main areas of self-directed learning of MIT OpenCourseWare newsletter subscribers, namely their: (1) learning preferences; (2) goals and motivations; (3) achievements; (4) obstacles and challenges; and (5) opportunities for life change.

**Learning preferences recap.** The survey results indicate that the paths and places for informal and self-directed learning are quite diverse. In the 21st century, people are engaging in self-directed learning when in cafes, airports, cars, and subways and trains, and even while hiking or walking, albeit less frequently than at home, work, school, or in a university.

In addition to the places and paths for self-directed online learning, the devices for informal learning are proliferating and becoming increasingly mobile. With the range of devices for accessing learning increasing, it is difficult for the designers of such learning to know how people might be accessing it. In addition, as the formats for delivery multiply, it places new demands on the designers of such instruction to offer potential learners multiple ways to find and use such information.

Not surprisingly, these learners are relying on open educational resources for their information needs. Most commonly, people tend to rely on common resources like Wikipedia and YouTube for their basic information needs. However, many other websites and resources were revealed as vital for informal learning online. Looking across these resources, it seems that online referenceware (e.g., Ask.com) is highly popular as are shared online video resources (e.g., How Stuff Works) and technology sites (e.g., Stack Overflow; a question and answer site for professional and enthusiast programmers). Online resources with the potential for life change included MIT OCW. Other life altering resources included the BBC online, iTunes University, the Khan Academy, TED talks, Coursera, edX, and Udacity along with Wikipedia and YouTube. What is interesting is that many of these resources are MOOC providers or OCW. The potential of open education such as OCW, therefore, is more than just acquiring new knowledge, skills, or resources; it can actually transform one’s life.

**Goals and motivations recap.** The respondents had many reasons for seeking information online. One factor that was apparent was that people long for internal self-improvement and professional development. In terms of specific subjects or disciplines, math and science skills as well as picking up a new language were the most popular. From a motivational standpoint, the respondents were curious or wanted to fill a specific information need, though a significant number of them wanted to learn something that could help society. As humanists like Carl Rogers, mentioned earlier, and many scholars from the self-directed learning movement would have predicted, a key aspect of using open educational contents is personal freedom to learn as well as choice, control, and fun. Many self-directed learners also long to produce or create something; and, as noted in the qualitative data results, for some individuals this might even be a small business.

The qualitative findings of this study spotlight the possibilities for self-directed learning through informal education channels, including OCW, OER, and MOOCs. Documenting life changes from informal learning of open ended content can hopefully serve to inspire others. The findings of this research offer new insights into the lives of individuals attempting to learn from free and open resources on the web. There are myriad reasons why individuals around the world access OER, OCW, MOOCs, and other free and open educational resources to learn. As such resources proliferate, so, too, does learner choice and autonomy. Stated another way, when the resource pool increases, there are new opportunities for self-directed learners to explore, share, and obtain new skills and competencies. Most notable are not specific skills in a certain subject area, but the opportunities for life altering changes. Better understanding of the goals of these self-directed learners should help in the design of more engaging informal and extreme learning environments. Such insights might also play a role in supporting people who presently are not self-directed learners.
**Achievements recap.** In terms of content related successes, our analyses of the data revealed that these learners amassed skills in physics, computer science, teaching, chemistry, business, law, and many other fields. However, there was no one resource that was pivotal to most learners; they learned through an assortment of online resources. In the open ended items, these individuals discussed their new sense of freedom to learn from a wide gamut of web-based tools and content resources. Many of our respondents truly enjoyed learning for the sake of learning, without any associated external reward or certificate of accomplishment. They explored personal interests and found what is meaningful to them.

**Challenges and obstacles recap.** Self-directed online learning is not without significant challenges today. Among the key obstacles or challenges that respondents faced were the lack of time to utilize the vast resources at their very fingertips. Given the explosion of open educational resources, that should not be a revelation. What was more surprising was that the second biggest obstacle was not related to issues of access or firewalls, but membership and technology costs or fees. We had anticipated that OCW subscribers would primarily utilize free and open online resources. Other researchers might follow up on this cost issue, while instructional designers and instructors might provide better access to high quality free resources. Naturally, there was a concern about the quality of the materials found and the lack of support to use them once they were deemed of high enough quality to use. To address such concerns, Mishra and Kanwar (in press) as well as Swan, Day, Bogle, and van Prooyen (in press) offer quality assurance criteria and advice for those tapping into MOOCs and open education.

**Possibility for life change recap.** While data analyses are ongoing, the vast majority of those that we surveyed felt a definite life change from free and open educational content. Given such results, we conclude that informal and self-directed learning, while not necessarily improving test scores, is playing a monumental role in society. As noted earlier, respondents wanted to produce or create something. For instance, some started a new business or learned marketing or web design skills to expand an existing one. Others learned skills that helped them to overcome huge problems such as when a country imploded or was dramatically altered politically (e.g., Russia). In contrast, several of the respondents noted that while they did not experience significant life changes, they were extremely happy to keep up with changes occurring around them. Others felt that changes were more incremental than suddenly life changing. Still others felt momentous life change in the form of newly discovered freedom to learn, autonomy, and a sense of Eureka as well as greater confidence, happiness, and inspiration. Their personal transformation was psychological not informational or intellectual. For most, open education offered a sense of accomplishment that was not possible in traditional educational settings.

What this research project reveals is that people have specific career-related or personal interests that they can now enjoy at their own leisure. The web has tremendously expanded access to education and provided learning options that previously did not exist. In the midst of these new information access points and novel learning routes, there is much internal or intrapsychological change occurring that, for the most part, is positively impacting one’s self-concept as a learner.

**Instructional implications**

There are several key implications from this research. Chief among these implications is that educators, instructional designers, and online learning tool and resource designers need to embed a sense of choice and control when creating or enhancing OCW, OER, and MOOCs. Some opportunity for personal fun and building or generating something is also important. Humans are increasingly migrating to online environments for their learning needs. They are curious and want to explore how stuff works. Often it is not course credit, a transcript mark, a certificate, or a degree that they are pursuing. Instead of such extrinsic motivators, they want to pursue their personal goals and passions. Furthermore, they not only want personal growth and professional development as a means to enhance their careers, but many also want a chance to help others pursue their learning interests. As such, these self-directed learning environments might embed online discussion forums and learning communities that can foster such knowledge support and caring.

Instructional designers might design instructional scaffolds and other such guidance to help learners select high quality of the online resources. Already, Class Central, TechnoDuet, OER Commons, the MOOC List, and Open Culture exist to help sort through the mass of educational contents now available. As learners as well as educators in
the coming years and decades come to embrace self-directed and informal learning paths, such guidance will be critical to learner efficiency and ultimate success. Self-directed learning perspectives and frameworks from Garrison (1997) as well as Song and Hill (2007) might help in that regard. In building on such efforts, Kop and Fournier (2010) conclude that in an age of MOOCs and OER, learners will require more sophisticated tools and resources to help them evaluate the viability of information available for their self-directed learning pursuits.

**Study limitations**

In the midst of these interesting results related to life change or impact, there were several key limitations that should be pointed out related to this particular study. First of all, the survey respondents were subscribers to a newsletter related to MIT OCW. It is likely that many or most such individuals have extensive self-directed learning experience from open educational contents, thereby limiting the generalizability of the findings. Along these same lines, it was not unexpected that subscribers to MIT OCW would have a preference toward math and science content; however, that undoubtedly does bias the results. Third, 44 percent of these respondents were from North America where access to the Internet is typically higher relative to most other countries and regions of the world. Fourth, one should also keep in mind that these MOOC participants volunteered or self-selected into the study. A fifth limitation was that of the 156,000 active subscribers to the MIT OCW newsletter, only 1,429, or less than 1 percent, responded to the survey; about half of which completed the open ended items. Finally, as a mixed methods study, only a portion of the extensive and highly interesting qualitative results could be included here due to length limitations.

**Final reflections and future directions**

It has been a little over a decade since Charles Vest’s courageous announcement about all courses from MIT being made available for free use online. The educational world has changed dramatically since that time. Not only are millions more people learning online in every educational sector—K-12, higher education, and corporate, military, and government training settings—but informal learning has simultaneously proliferated. The movement toward a more open educational system has shifted to highly massive endeavors that are prominent in the news such as MOOCs. As open education rises in salience, most institutions of higher learning are deliberating on next steps. Some are struggling to come up with plans and solutions that incorporate open education contents and courses and nurture self-directed learning on an ever increasing scale. At the same time, for economic, political, and social reasons, some states and universities are quickly sponsoring MOOCs and other related open education initiatives.

With educational institutions and organizations shifting resources toward more free and open contents, there is a pressing need to understand how to foster productive self-directed learning from OER. First the characteristics of self-directed learners and processes of self-directed learning need to be better understood. Second, as these traits are uncovered, training programs on self-directed learning might be designed. In addition, self-directed learning supports or scaffolds might be embedded in OER or MOOCs at key moments in the learning process or made available upon demand. And, as indicated earlier, professional communities of self-directed online learners might be formed and supported.

Without a doubt, as with President Kennedy and his dreams to explore space four decades earlier, Charles Vest ushered in this new age of open education and self-directed online learning with his bold announcement back on April 4, 2001. Unlike Kennedy, Vest fortunately lived to witness firsthand the substantial fruits of his ideas before passing away on December 12, 2013 (Brant, 2013). By that time, thousands of MIT open courses had been accessed by more than 150 million learners around the globe. Moreover, the popularity of OCW has inspired hundreds of additional colleges and universities to enter the OCW movement, establish e-learning programs, and experiment with MOOCs and other types of open courses.

Vest certainly will not be the last leader or figurehead of the open education movement, but with one small step, he and others at MIT took the first gigantic leap for the free and open education of humankind. Back in 2001, we no longer needed challenging goals related to exploring outer space as in Kennedy’s days; instead, Vest wisely envisioned how traveling within a virtual space could open up the entire educational world to places never previously experienced or touched. As the findings of this study testify, that mission was definitely accomplished.
Acknowledgements

We would like to thank Steve Carson, former External Relations Director of the MIT OpenCourseWare (OCW) initiative, as well as Yvonne Ng and Cecilia d'Oliveira from the MIT OCW project for their support of this project.

References


The Multimedia-Based Learning System Improved Cognitive Skills and Motivation of Disabled Children with a Very High Rate

Sawsan Saad*, Amal Dandashi, Jihad M. Aljaam and Moataz Saleh
Department of Computer Science, Qatar University, Doha, Qatar // s.saad@qu.edu.qa // amaldandashi@qu.edu.qa // jaam@qu.edu.qa // moataz@qu.edu.qa

*Corresponding author

(Submitted February 25, 2014; Revised June 20, 2014; Accepted July 24, 2014)

ABSTRACT

A multimedia-based learning system to teach children with intellectual disabilities (ID) the basic living and science concepts is proposed. The tutorials’ development is pedagogically based on Mayer’s Cognitive Theory of Multimedia Learning combined with Skinner’s Operant Conditioning Model. Two types of tutorials are proposed. In the first type; the contents are developed in a static manner while in the second type; the contents are developed using a domain-specific ontology with text processing tools and Google search engine. Both types of tutorials have shown a considerable improvement in the learning process and allowed the children with ID to improve their cognitive skills and become more proactive in the classroom.

Keywords
Multimedia Tutorials, Learning Model, Intellectual Disability, Ontology

Introduction

Children with intellectual disabilities (ID) are very challenging to teach, and can learn best with the use of different methodologies that engage their senses, such as using images, sounds and clips. In fact, every child in the same classroom may have a different way of understanding, and should have an individual learning plan with specific goals and objectives. Therefore, instructors use different methods to teach these children in an acceptable manner (Adam & Tatnall, 2008).

Existing research in Universal Accessibility (UA) related to individuals with ID is quite limited. Studies that examine cognitive disabilities started to emerge only in the past decade. Cohone et al. (2007) worked on developing reminiscence tools for people with Alzheimer’s disease. Wu et al. (2007) developed some tools for people with amnesia. Moffatt et al. (2004) developed tools for people with aphasia. A field that has drawn increasing attention both in the general public and in academia is the study of children with autism and the attempt to improve their language development and social skills via computer mediated software or agents e.g., Lehman (1998), Hart (2005), and Tartaro & Cassell (2008). In addition to developing technologies to be used by children with autism themselves, researchers also explored technologies to assist the children’s caregivers in communication, record collection and analysis, decision-making, and assessment of the children’s internal states as what was done by Kientz et al. (2007). A few studies investigated people with ID together with users of various other cognitive impairments e.g., Dawe (2006). These studies provide valuable insights regarding the impact of cognitive impairments on the use of computer-related activities. But due to fundamental differences between conditions such as Alzheimer’s disease, amnesia, aphasia, Autism, and ID, observations and findings from these studies cannot be readily applied to children with ID.

Researchers in special education are to some extent informed of the potential of computer technology in helping individuals with ID, e.g., Buckley (2000) and Black & Wood (2003). They claim that computer technology can help people with ID increase confidence and motivation through creative activities and web browsing. Computer technology also has other benefits, including errorless learning, patient feedback, immediate feedback, self-paced learning, and independence of learning. However, Lloyd et al. (2006) suggested that the actual benefits of computer technology may be reduced or not apparent depending on the quality of the software. First, the contents of many software programs are not age appropriate. Second, many educational software are unable to reach educational goals and are used as a tool for mere entertainment. Third, many applications do not promote independent learning.

Different approaches have been proposed to present the multimedia-based learning system (Adam & Tatnall, 2008; Evans et al., 2006; Garcia-Ruiz et al., 2008; Kirk et al., 2011). The empirical study conducted by Ortega-Tudela &
Gómez-Ariz (2006) examined the impact of educational software on learning mathematical counting skills. Eighteen children with ID have participated in the study. Ten of them used multimedia education software to learn basic counting skills and the other eight tried to learn the same counting skills via the traditional paper-and-pencil approach. After fifteen sessions, children who used the educational software demonstrated significantly higher performance than those who used paper and pencil. Wuang et al. (2011) illustrated a system that applies design and learning theories in developing multimedia courseware. This system develops multimedia courseware of Loci in two dimensions (Li2D) using ADDIE methodology with Macromedia Flash 8 and Adobe Photoshop. The theories that have been applied in the development include: Design Theory, Behaviorism, Constructivism, Cognitive and Van Hiele Thinking Model. A Computer Reinforced Online Multimedia Education (Crome) framework was proposed in (Dong & Li, 2006). The authors presented a method to integrate main components of learning, teaching, testing and adaptive student modeling. Crome was designed to incorporate portability, reusability, scalability and interoperability. Cheng et al. (2009), proposed a multi-ontology based multimedia annotation model in order to ensure effective utilization of multimedia by different users. In this model, domain independent multimedia ontology, based on MPEG-7 content description tools, was integrated with multiple domain ontologies to provide multiple domain-specific views of multimedia content. A term extraction procedure was designed to automatically extract domain specific ontological terms from textual sources. Evaluation results prove that multi-ontology based multimedia annotation enhances various users’ information needs.

Although there is little research on computer usage of individuals with ID, there is one well-documented design case for this population. The National Down Syndrome Society (NDSS, 2009) worked with a Web design firm to develop a Web site specifically designed for people with ID, called Web Fun Central. The goal was to teach Web browsing and other computer skills specifically. Six individuals with ID became part of the design team and took part in usability testing sessions. Two separate usability tests, both involving the same six participants, were completed to inform the design. As a part of this process, the researchers were able to develop a number of guidelines for designing Web sites and computer-based learning modules for people with ID.

In this work, we propose an assistive educational system using multimedia technology to teach the children with ID the basic concepts of daily living tasks and elementary sciences in an attractive way. Such a system is highly needed due to the lack of appropriate contents for children with ID in the state of Qatar. Our proposal generates the multimedia tutorials which can be easily used by the instructor and provide mentally disabled children the opportunity to learn with tutorials designed to suit their intellectual needs. We use a combination model of two cognitive theories, as is elaborated in section 2. The development of the tutorials is based on two techniques; the static manner is presented in section 3, and the dynamic technique in section 4. Assessment and analysis is presented in section 5. The last section concludes the paper.

Learning models

Both types of proposed tutorials, dynamic and static are developed with a combination view of two cognitive models: Mayer’s Cognitive Theory of Multimedia Learning which implements using multimedia elements to enhance the learning experience of a child, and a mild implementation of Skinner’s Operant Conditioning which suggests using gentle positive and negative reinforcements to help motivate the child.

Mayer’s cognitive theory of multimedia learning

Mayer’s Cognitive Theory of Multimedia Learning (Mayer & Alexander, 2011) allows children to use their auditory and visual channels in the learning process. It involves active use of their sensory, working and long-term memory to process multimedia elements into logical mental constructs. This theory assumes the following:

- There are two main channels for processing information; auditory and visual.
- Each channel has a finite capacity for cognitive load.
- Filtering, selecting, organizing and integrating information is an active part of the learning process.

According to Mayer, there are three important cognitive processes, which the multimedia learner engages in. The first one involves selecting verbal and visual information to yield a learning base, the second involves organizing verbal and visual information to form into coherent mental representations, and the third one includes integrating the
resulting verbal and visual representations with one another. The following figure represents this cognitive process (figure 1).

![Cognitive process of multimedia model](image)

**Figure 1.** Cognitive process of multimedia model

**Skinner’s behaviorist operant conditioning model**

Skinner’s Behaviorist Operant Conditioning Model (Staddon & Cerutti, 2003) is a process that encourages behavior through positive or negative reinforcement. Reinforcement may come in the following forms:

- **Positive reinforcement**: favorable event given to a child after an achievement (praise, reward, etc.).
- **Negative reinforcement**: the removal of an undesired outcome after a positive achievement is made by the child.
- **Positive punishment**: unfavorable event is given in order to weaken the following response.
- **Negative punishment**: favorable event is removed after undesired behavior occurs.

According to Skinner, there are three types of responses that can alter behavior:

- **Neutral operants**, which are responses from the environment or other unaffected factors, which neither increase nor decrease the probability of a certain behavior happening.
- **Reinforcers**, which are responses that increase the probability of a certain behavior being repeated. These reinforcers may be positive or negative.
- **Punishers**, which are responses that decrease the likelihood of a certain behavior being repeated.

While Skinner’s theory implies that human behavior is solely affected by external factors such as what the subject is being exposed to, we note that external factors merely play a role in the child psychology and learning process, and the ideology of Behaviorism may be integrated with other techniques which do take into account internal factors such as thinking, emotions and previous experience. For the purpose of this study, only the positive and negative reinforcement will be mildly used, and this is in order to promote positive encouragement of the child rather than negative.

**Static technique**

The static technique consists of setting first the lessons objectives and contents with the assistance of the special education instructors. Our system supports and monitors the learning process through providing several interactive educational multimedia contents consisting of text, images, and short clips. These multimedia elements are linked to the objectives of teaching personalized plan. The system can be used in two ways:

- **Group activities**: The system uses a smart board to view multimedia contents, i.e. lessons, puzzles, quizzes, etc. The children can interact with these contents.
- **Individual activities**: each child has his own tablet. The teacher controls the class by sending different multimedia contents to the children’s tablets according to their level of abilities and personalized teaching plan. The children are able to practice the contents at home as the teacher will daily transfer a copy of these contents to the children’s tablets.
The users of the system

The system has four types, which are: administrator, teacher, student, and parent.

- **Administrator:** It is the responsibility of the administrator to manage all users’ accounts including adding deleting, and updating. He/she uploads the multimedia contents and the plans. He/She links these contents with their corresponding objectives. System settings like sending parents periodic SMS/Email for feedback on their children progress are also handled by the administrator.

- **Teacher:** The learning process in the class is managed by the teacher who can control the educational multimedia contents to be viewed on the smart board as well as the students PC tablets. The teacher also transfers the daily covered contents of each student to his/her tablet for home revision with the parents.

- **Student:** This is the main user of the system who can smoothly interact with the contents available on both class smart board and PC tablet. Upon completion of a lesson, the system provides the student with several puzzles and games to evaluate his/her understanding of that lesson. This process is done with multiple trials, and intelligent algorithms are built to indicate student’s mistakes. The system tracks all student results, records it in database and eventually reports the student performance.

- **Parent:** The system effectively contributes in increasing the level of interaction between the parents and their children through: First, sending periodic short SMS and Emails about the children progress. Second, ability of accessing the system website to post feedbacks about their children personalized plan. Third, having the class materials installed on the children PC tablet gives the parent the opportunity to follow up with their children’s daily classes. Finally, parents can review their children’s daily lessons and contact teachers for any inquiries.

Educational content

The system provides a large number of contents to meet the objectives of the learning process for the children. These contents are designed and customized based on the international educational system FACE curriculum, which is widely used to teach children with special needs and it covers different scopes including math, science, reading, writing, religion, and social life. The contents are designed with a focus on:

- Represent the local Qatari environment such as dress, food, shops and currency.
- Suit students’ intellectual, vision, and hearing capabilities (i.e., levels of difficulty, sounds, and Developed color).
- Interacts easily with students to achieve the learning objectives.
- Motivates students through employing exciting contents as opposed to rigid teaching.
- Varies in teaching styles (i.e. multiple choices, match objects, drag and drop, find similarities, and puzzles).
- Features the ability to repeatedly update, improve, and personalize contents. Organized and sequenced logically i.e. from easy to difficult, and basic to advanced.
- Organized and sequenced logically i.e. from basic to advanced.

Next the characters and the different scenarios are reviewed by special education instructors in order to achieve high quality tutorials. Then, the lessons are designed with the help of a graphic designer and multimedia developer. Finally, we add the animations and the corresponding sounds using the standard Arabic language, which are recorded in a studio. Final tutorials are stored in a specific database and can be used incrementally according to the children’s performances. Every tutorial needs approximately three months in average to be completed properly. Our multimedia tutorials are developed using Adobe Flash CS3, and Action Script. They consist of high quality animations at a video frame rate of 32 frames per minute. All lessons are taught in Arabic with the use of multimedia tools that appear attractive to children. Figure 2 shows some snapshots from the static tutorials.

![Figure 2. Snapshots from the static tutorial content](image)
Dynamic technique

In the second technique, we propose to generate multimedia tutorials dynamically through providing an automatic mechanism to query for a desired topic based on semantic content analysis and ontology. The instructor can easily customize the tutorials based on the specific needs of each child in the classroom. The dynamic tutorials would become more time efficient due to the machine learning process of the automated system. The system will statistically learn the preferred customization per student and will generate automatically customized tutorials.

Figure 3 gives an overview of the different steps of our proposed system. The first input is the educational text that will be processed using text-processing tools (i.e., segmentation, stemming, analyzing). We parse text with the Stanford Parser that aids to retrieve sentences containing pairs of keyword entities and represents them by using a data structure called a “Dependency Graph” (Klein & Manning, 2003). The objective is to identify the actors (i.e., lion, elephant, tiger), the objects (i.e., water, plant), and the actions (i.e., attack, live, eat) in the story text. Each retrieved sentence is analyzed to extract the raw relationships between keywords entities. The extracted relationships and entities will be semantically mapped and validated according to the domain description captured in an ontology model. All raw relationship instances are sent as semantic queries in order to retrieve the available multimedia instances in our ontology (i.e., the knowledgebase). Complimentary queries will sent also to the Google image database through Google APIs to fetch the corresponding multimedia elements. Hence, the teachers can then select the appropriate multimedia elements that are suitable to customize the tutorial based on the needs of each child in the classroom.

**Figure 3.** An overview of educational text analysis method

Educational Text Parsing and Semantic Relations Extraction

In this step, the text input is segmented into sentences. Each sentence will be processed syntactically to recognize its part of speech. The goal is to provide the syntactical structure of educational text sentences. The word that belongs to a verb or a noun grammar category is defined as an entity. We focus on sentences that mention at least two keywords entities. The parser returns “parse trees” which is a rooted tree. Each node in the tree represents a part of the sentence, as in Figure 4. The parse trees generate by using the Stanford Parser that is a statistical natural language parser (Sleator & Temperley, 1995). It uses a set grammatical function of words to record the most likely syntactic structure of a sentence.

We next transformed the parse tree into a Dependency Graph (DG) format (De Marneffe et al., 2006) that we analyze to extract the semantic relationships. The DG captures the implicit dependencies (semantics) between keywords and entities in a sentence. Each node is a word and labeled edges represent grammatical relations between words. DGs are rooted, oriented, and labeled graphs that make them easier to read and process than parse trees or other representations.
Therefore, we extract the semantic relations in a subject-predicate-object triplet form, corresponding to the RDF statements (Klyne & Carroll, 2006). Subjects and objects represent words entities. Predicates are used to describe the nature of the relationship between two entities with a sequence of words. In order to realize a good extraction, we use a set of grammatical patterns that express relations between entities. These deep dependencies linguistic analysis of sentences will be perform by using the grammatical formalism called link grammar (Sleator & Temperley, 1995). In this formalism, the links are labeled depending on the nature of the grammatical relationship of two terms within the sentence. For example, the link label "S" is used to connect a subject to a verb, while "D" is used to connect a determiner to a noun. If the terms are not directly connected, one of the properties of the formalism, called connectivity, ensures that all terms of the sentence are at least indirectly connected via a number of intermediary terms. A path between two words of a sentence is called a link-path. The source and target of a link-path are denoted as start term and end term respectively. The set of all links describes the grammar of the entire sentence is referred to as linkage. The sequence of words interlinked on that path is called word-path and it can be seen as the semantic relationship between two terms.

Given the sentence "Crocodile is an aquatic animal".

- Start-term: “Crocodile”
- Stop-term: “Animal”
- Word-path: “is,” “an”
- Link-path: “Ss, Ost, Ds, A”

Figure 5 illustrates the linkage of the example sentence. The connected terms used in the example is highlighted. The word-path describes the relation between start and stop term. By combining the terms in the word-path to form a single predicate, the valid semantic relation Is-a (crocodile, animal) is generated.
For finding a set of valid link-paths, we propose manually annotating a set of semantic relation triplets. The table shows grammatically distinct types of predicates can be found by using predefined link-paths. The link-path such as (LiveIn) uses the information given by two terms that function as subject and object to a verb. We can use it as predicate LiveIn(crocodile, tropics). While in the case of the link-path (is-a), we have two nouns which are connected with a predicate, as we have seen is the previous example of Figure 5.

<table>
<thead>
<tr>
<th>Link-path</th>
<th>Predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ss Ost Ds</td>
<td>Is-a</td>
</tr>
<tr>
<td>Ss Ds</td>
<td>LiveIn</td>
</tr>
<tr>
<td>Ss PvMvJp</td>
<td>WasAttackedBy</td>
</tr>
</tbody>
</table>

**Ontology learning development**

Since ontologies provide a shared understanding of a domain of interest, they have become a key technology for semantics knowledge extraction and integration. The aim of ontology is to develop knowledge representation that can be shared and reused. Guber (Gruber, 1995) defined ontology as: “A formal explicit specification of a shared conceptualization.” Domain ontology provides particular meaning of terms as they apply to that domain. In our system, the use of ontology can provide a common underlying language that aids to understand the extracted knowledge from different educational texts. Ontology can help to provide semantic concepts that we can use for searching and retrieving the multimedia elements to build the tutorials.

Therefore, we write down in an unstructured list all relevant terms that are expected to appear in the ontology and we would like to provide an index of specific children’s tutorial (animal, food, etc.). For animal ontology as an example, we could make the following questions to extract the most important animal-related terms. Where they live? What they eat? How dangerous they are? And a bit about the basic anatomy (number of legs, wings, toes, etc.). Therefore, the terms list will be include: animal, subtypes of animal (herbivore, carnivore), different names of animal (i.e., crocodile, elephant, tiger, bird, etc.), location, plant, water, animal actions (i.e. eat, attack, etc.), and so on.

For developing the concept hierarchy, we usually start with the definition of the most general concepts in the domain and subsequent specialization of the concepts. From the list created previously, we select a group of animal-related terms and ask what they have in common and what “siblings” there might be. Hence, we specialize the animal class by creating some of its subclasses such as (herbivore, carnivore). We can further categorize the carnivore class into (crocodile, tiger, lion), and so on. These concepts will be arranged in a hierarchical taxonomy to form the ontology. All the siblings in the concept hierarchy must be at the same level of generality. The concepts alone will not provide enough information to answer all the questions about the animals. Once we have defined some of the concepts, we must describe the internal structure of concepts. For example, Animal parentOf animal, elephants eat planets.

The underlying formalism for ontology is Description Logics (DLs) (Baader & Nutt, 2003), which are a family of knowledge representation formalisms that have formal semantics. This family of logics is tailored towards representing terminological knowledge of an application domain in a structured and formally way. Description logics allow users to define important notions, such as classes or relations of their application domain in terms of concepts and roles. These concepts (unary predicates) and roles (binary predicates) then restrict the way these classes and relations are interpreted. Based on these definitions, implicitly captured knowledge can be inferred from the given descriptions of concepts and roles. These inferences are defined based on the formal semantics of DLs.

The central notion for DLs is concept descriptions, which can be built from concept names. For example, one can describe animal ontology in DLs as an Herbivore that eats only Plants as follow:

\[
\text{Herbivore} = \text{Animal} \land \forall \text{eats. Plant}
\]

The last step is creating individual instances of classes in the hierarchy. We use Protégé editor(Protégé, 2007), which is a free, open source ontology editor and a knowledge acquisition system. It is a tool supporting the construction of ontologies and it also provides an application platform for knowledge-based systems. Defining an individual instance of a class requires: (1) choosing a class, (2) creating an individual instance of that class, and (3) filling in the values.
After we define an initial version of the ontology, we can evaluate and debug it by using it in applications or by discussing it with experts in the field, or both. As a result, we will almost certainly need to revise the initial ontology. This process of iterative design will likely continue through the entire life cycle of the ontology.

**Multimedia tutorial building**

Our system applies the mapping process between dependency graphs and our ontologies according to each topic. The extracted relationships and entities will be semantically mapped and validated according to the domain description captured in an ontology model. Thus, we need to leverage the degree of the similarity between text-extracted entities that correspond to the nodes in the dependency graph and the concepts in the ontology (called concept entity). Once, we find an instance match for the subject and the object. We search to find a matching property for the predicate in the domain ontology model. Finally, we check if the concepts to which the instances for subject and predicate are asserted in domain and range of the property matched.

After determining the mapping concepts, we can combine them into semantic queries in order to retrieve the corresponding multimedia instances. A query language such as SPARQL seems to be best appropriate tool (http://www.w3.org/TR/rdf-sparql-query). SPARQL is particularly adequate for extracting data from ontology and through its construct statement, we can generate new data. The SPARQL query is executed against the knowledge base, which returns a list of instances that satisfy the query. Then the multimedia instances that are retrieved, ranked, and presented to the instructor. Complimentary queries will sent also to the Google image and video database through Google APIs to fetch more multimedia elements. Finally, the instructor can then select the appropriate multimedia elements that are suitable to the needs of the children in the classroom. Teachers can get additional multimedia elements either from the ontology or from Google search engine by simply clicking on the words or sentences of the text. Figure 7 shows the retrieved multimedia elements corresponding to the following educational text: “Crocodile is a aquatic animal that lives in the tropics in Africa. Crocodiles attack the elephants.”

![Figure 7. Multimedia elements associated with the educational texts](image)

**Assessments**

The proposed learning systems have been tested on 100 children with ID from the Shafallah Center in Doha. The children have an average mental age of 8 years, and are mildly disabled. Half of these children have Down syndrome.
(DS), and the other half have various intellectual disabilities (ID). The selected children have some basic ability of reading simple Arabic words, doing simple calculations, recognizing and pointing to objects. Twenty special education instructors from the Shafallah Center participated in the assessment to assist the children using in the learning sessions. A training session was conducted for these teachers to explain to them how to use the system along with the tutorials and the exercises.

The children are first allocated to try the static multimedia tutorials, and are subjected to corresponding exercises that indicate their comprehensive understanding level of the lessons. Their performance scores are assessed by their usual instructors according to an assessment model developed for this testing session. Next, the children are allocated to test the dynamically generated multimedia tutorials on a different set of lessons, but in a similar level of difficulty. The dynamic method involves the instructors teaching the lessons in the conventional story book method, and then reinforcing the concepts by dynamically extracting multimedia elements that correspond to the lesson, using the developed system. The children’s knowledge is then tested using a set of exercises, and their performance scores are assessed once more by their instructors by using a pre-developed assessment model. In addition to gathering performance score information, we have developed an assessment rubric to measure motivational levels of the children. The motivation assessment rubric consists of scoring the relative motivation level of each child, as observed by its regular special needs instructors. The scoring method involves a motivation spectrum measure of 1-10, 1 being least motivated, and 10 being most motivated. As an average in Shafallah Center, each group of 5-6 children have one special needs instructor dedicated to their class. Therefore we were able to rely on the special needs instructor to score the relative motivation levels as observed while using multimedia tutorials, static and dynamic.

The children are assessed based on a scoring system that takes into account the following constraints: physical age, mental age, disability, gender, timing, number of correct answers, number of repetitions required to understand the lesson, mood and motivation of child, and exceptional hardships. The obtained results indicate the following about both dynamic and static models: increased motivation, improved performance, relatively speedy learning and improved memory function and memorization.

The assessment process is divided into three parts: static tutorial assessment, dynamic tutorial assessment and combination assessment.

**Static Tutorial Assessment**

Static assessment consists of evaluating the average performance scores, timing, and motivation levels. Results are differentiated between children with DS and other IDs, and females versus males. Figure 8 depicts results in terms of percentile scores, relative timing and motivation levels observed, comparing the results of the children with ID with the children with other DS. Note that the time depicted here represents the average time (in minutes) that each student took to complete an exercise. As shown in the graph, we can see that children with ID and DS both scored high average scores ranging from 70% to 80%, although we note that the children with DS slightly outperform the ones with other IDs. High motivation levels for both groups were also reported with an average of 85% to 90% children exhibiting increased motivation. We can conclude that the implemented multimedia lessons have been equally useful and effective for all groups of disabilities.

![Figure 8. Performance of ID vs DS – Static tutorials](image)
Figure 9 illustrates the performance of females versus males, without taking into consideration differences in disabilities. As shown in the graph, female children have outperformed the male children by a 15% range. The average score of all females is 81%, while the average score for males is 62%. It is worthwhile to note that while the females fared better in terms of scores, the males took less time to complete the exercises, which is indicative of the relative nature of the genders; females tend to be more patient and objective oriented, while males tend to be impatient and view the lessons and exercises as games rather than lesson scores. We can conclude from this graph that the multimedia lessons had better results on females rather than males.

**Figure 9. Female vs Male performance – Static tutorials**

**Dynamic tutorial assessment**

Dynamic assessment, similarly to the static assessment, consists of evaluating the average performance scores, timing and motivation levels. Results differentiate between children with DS, ID, females and males. Figure 10 demonstrates results in terms of percentile scores, relative time and motivation levels. Children with DS slightly over performed the children with other ID, scoring within a higher range of 10%. Children with other IDs were also noted to have taken a longer amount of time to complete exercises. Both groups exhibited high motivation levels, although the children who scored better (DS) are more likely to have higher motivation to keep trying.

**Figure 10. Performance of ID vs DS – Dynamic tutorials**

Figure 11 depicts the performance results of females versus males, regardless of disability differences. We can observe from the graph females outperformed males within a 15% margin, but similarly to the static tutorial results, we notice that females take more time to answer exercises. Dynamic tutorials are more difficult to precisely assess given the time constraint, due to the process of the student first studying the concept in conventional classroom methods, then using dynamic tutorials to back it up. Some students had some difficulties tying the classroom concept with the multimedia concept, as this method is still novel to them. Females exhibited higher motivation levels than males.
Combined assessment

The combined assessment consists of the evaluation of both static and dynamic systems in combination and comparison to each other. Figure 12 depicts the acceptance and relative performance levels of all the children. 7.5% of the children tested with the static tutorials refused to accept or participate in the lessons, due to some physical stress they received from the sounds associated with the lessons. 2.5% participated but found it too difficult. For the dynamic tutorials, we have a slightly higher acceptance rate, as only 5% of students rejected participation and another 5% found the concept of linking the lessons with the multimedia elements too difficult. Relative performance levels significantly increased, when compared to the performance of the children at conventional lessons, without the use of multimedia. 88% of the children exhibited higher relative performance with the static tutorials, and 82% had higher performance levels with the dynamic tutorials.

While statically generated tutorials have proven to be a good source of education for the children with IDs, allowing for increased cognitive levels, motivation, enthusiasm and high relative performance levels, the problem faced with these pre-designed tutorials is the time consumed to generate each. Each tutorial, designed as per the specific curriculum specifications for each class or school, requires one month of work by animation, script and sound studio experts. Whereas, the benefit behind dynamically generated tutorials is to be able to generate a customized tutorial on the spot, in the classroom, with no predesigning requirements. Furthermore, the instructors can chose multimedia elements to complement the lessons and can customize each tutorial combination for each student as seen fit. Testing with the dynamic tutorial generation has proven to be similarly as effective for the children with ID, has increased cognitive levels, motivation levels and relative performance levels. Figure 13 illustrates the average timing required by the instructors to generate each dynamic customized tutorial in the class time. As shown in the graph, 21% of trails required 6 minutes to generate a tutorial, 47% took 10 minutes, and the remaining needed an average of 15 minutes or more.
Table 2 below demonstrates a summary of average score, motivation and timing levels achieved by the students per static and dynamic tutorials.

<table>
<thead>
<tr>
<th>ID</th>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score: 74%</td>
<td>Score: 69%</td>
</tr>
<tr>
<td></td>
<td>Motivation: 86%</td>
<td>Motivation: 80%</td>
</tr>
<tr>
<td></td>
<td>Time: 6 min</td>
<td>Time: 8 min</td>
</tr>
<tr>
<td>DS</td>
<td>Score: 82%</td>
<td>Score: 75%</td>
</tr>
<tr>
<td></td>
<td>Motivation: 89%</td>
<td>Motivation: 86%</td>
</tr>
<tr>
<td></td>
<td>Time: 5 min</td>
<td>Time: 6 min</td>
</tr>
<tr>
<td>Female</td>
<td>Score: 79%</td>
<td>Score: 81%</td>
</tr>
<tr>
<td></td>
<td>Motivation: 90%</td>
<td>Motivation: 90%</td>
</tr>
<tr>
<td></td>
<td>Time: 7 min</td>
<td>Time: 9 min</td>
</tr>
<tr>
<td>Male</td>
<td>Score: 77%</td>
<td>Score: 63%</td>
</tr>
<tr>
<td></td>
<td>Motivation: 85%</td>
<td>Motivation: 76%</td>
</tr>
<tr>
<td></td>
<td>Time: 4 min</td>
<td>Time: 5 min</td>
</tr>
</tbody>
</table>

In an overview of the assessment results, it can be noted that children with DS seemed to achieve better results than that of children with other IDs, and females, while taking more time achieve better scores due to their detail oriented nature, whereas the boys tend to deal with the multimedia tutorials in a more playful manner. Nonetheless, all groups obtained high scores, and teachers noted an 80% - 88% increased relative performance in their students after having used either the static or dynamic tutorials. Instructors were enthusiastic about using the system in the classroom as part of daily lessons and noted the numerous benefits on the children’s part; including higher motivation levels, increased performance, faster cognition and high acceptance rates.

The limitation of our assessment was not testing out the differences in performance between static and dynamic tutorials, over prolonged periods of time. We were allocated one week of testing for the static tutorials, and one week of testing for the dynamic tutorials. The reason the time factor is so important because we hypothesize that the longer dynamic tutorials are used, the more effective they will be. However, being tested only over a period of one week does not allow us to gather a complete picture. The dynamic tutorials have to be used in accordance with conventional classroom methods, so the instructor would first introduce a concept in class, and then enforce the understanding of the concept with related tutorials. As a first trial for the children, some students faced some difficulties tying the two classroom concept with the multimedia concept. We predict that with further use results will be far improved. On the other hand, this assessment had various strengths: the number of children allocated to testing (100), having both genders tested, and having several types of IDs allows us to further understand how to target each group. Having the cooperation and dedication of the special need instructors of Shafallah Center was also an incredible strength to this assessment, as it could not have been done without their expertise.
Conclusions

This work proposed a novel assistive educational system to generate multimedia-based tutorials under a combinational pedagogical model: Mayer’s Cognitive Theory of Multimedia Learning and a mild implementation of Skinner’s Operant Conditioning model. The system is implemented in two main methods: (1) pre-designed or “Static Tutorials” where which tutorials are custom designed by animation experts according to specific curriculum needs and (2) “Dynamic Tutorials,” where which the following techniques are used to generate customized tutorials: text processing, relationship extraction, ontology building and dynamic extraction of online multimedia elements. The dynamic tutorials system allows instructors to generate tutorials on the fly, in the classroom, and in accordance to the lesson they are teaching. Instructors can choose multimedia elements appropriate for the specific needs of the child they are teaching, within minutes. The noted advantage of using the dynamically generated tutorial system over the statically pre-designed ones, is the amount of time and efficiency saved. The pre-designed tutorials require detailed lesson design by experts, and each tutorial requires up to one month of work. Whereas the dynamic tutorial system, once developed, requires minutes of time for the instructor to generate, and is far more flexible and practical, as it can incorporate many different lessons and does not have pertain to a set curriculum.

From the performance aspect, it has been observed that both static and dynamic tutorials had excellent results on the cognitive ability of the children with ID, allowing them enhanced learning, higher relative performance, higher scores, acceptance and motivation levels. The children exhibited an average score between 70% - 80% and even higher motivation levels ranging from 80% - 90%. In conclusion, the use of multimedia-enhanced learning has been found to be extremely beneficial for children with ID, as well as practical and efficient for their instructors.

Acknowledgments

This publication was made possible by a grant from the Qatar National Research Fund under its award NPRP 09-052-5-003. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Qatar National Research Fund.

References


Revisiting the Blended Learning Literature: Using a Complex Adaptive Systems Framework

Yuping Wang, Xibin Han* and Juan Yang
School of Languages and Linguistics, Griffith University, Nathan QLD 4111, Australia // Institute of Education, Tsinghua University, Rm412, Wennan Building, Qinghua Yuan, Beijing 100083, China // Institute of Education, Tsinghua University, 3 floor, North Building of Architecture, Qinghua Yuan, Beijing 100083, China // y.wang@griffith.edu.au // hanxb@mail.tsinghua.edu.cn // juan-yang@mail.tsinghua.edu.cn

*Corresponding author

(Submitted June 13, 2014; Revised September 19, 2014; Accepted October 23, 2014)

ABSTRACT
This research has two aims: (1) to bridge a gap in blended learning research — the lack of a systems approach to the understanding of blended learning research and practice, and (2) to promote a more comprehensive understanding of what has been achieved and what needs to be achieved in blended learning research and practice. To achieve these aims, we first assess the strengths and limitations in existing models of blended learning, then propose a framework for blended learning that is grounded in the complex adaptive systems theory. The proposed framework sees blended learning as a system consisting of six essential subsystems, and all the subsystems relate to and interact upon one another. The proposed framework is then applied to the review of 87 empirical studies from the current blended learning literature. The review identifies several gaps in current blended learning research and practice, and advances our understanding of some untapped potential of this new system of learning. We hope that this research will shed light on critical issues in understanding blended learning and in scaling up its implementation in tertiary education.

Keywords
Blended learning, Complex adaptive system, Blended learning model, Framework, Flipped classroom

Introduction

Because the term “blended learning” has been commonly used in tertiary education for well over a decade, it is perhaps not an exaggeration to say that it has become part of the vernacular. Blended learning has been gaining in importance, especially during the last five years, with the development of online learning and the recent rise of MOOCs and flipped teaching. Our literature review indicates that its effectiveness and validity as a new form of learning has been established in practice. At the same time, our review also reveals that the great majority of the empirical studies into blended learning are research interventions of short duration conducted at either the course or task level, focusing on just one or a few aspects of blended learning. As a result, investigations into blended learning continue to be fragmented and many important issues remain unexplored. This is highlighted by Owston (2013, p. 1), who stated: “There is a need for research investigating why blended learning, despite its many inherent advantages, has not been scaled up successfully in very many institutions.”

The term “blended learning” has been used interchangeably with “mixed mode learning,” “hybrid instruction,” and “technology-mediated/enhanced learning.” It has been defined and redefined by various studies, but none has provided us with a complete view of what constitutes blended learning and how different components of blended learning work together over time to achieve an integrated whole. Perhaps the most widely held understanding of blended learning is that it is a combination of “face-to-face instruction and computer-mediated instruction” (Graham, 2006, p. 5). This current study aims to promote a deeper understanding of blended learning research and practice, first using a different perspective—the complex adaptive systems perspective—and secondly, through a review of the recent literature on blended learning.

To achieve this aim, this article first assesses the strengths and limitations of existing blended learning models then discusses the theories of complex adaptive systems in an effort to develop a framework that effectively captures the nature and dynamics of blended learning. This discussion then leads to the proposal of a framework for complex adaptive blended learning systems, called the CABLS framework. We then apply this framework to a review of the recent blended learning literature to identify gaps in current blended learning research and practice. We hope that this...
research will promote a more comprehensive understanding of what has been achieved and what needs to be achieved in blended learning, in terms of both research and practice.

A review of blended learning models

During the last 15 years, a great number of blended learning frameworks and models have emerged, and these have advanced our understanding in many important ways. The following review discusses a few of the most influential models, and this discussion provides diverse lenses through which to view the differences between these models and the one proposed in this research.

Shea’s grounded model promotes a pyramid framework starting with “our assumptions and beliefs about the nature of knowledge” (2007, p. 31). This is followed by the identification of “the theories of learning that reflect these philosophical underpinnings,” the articulation of “complementary pedagogical approaches,” “instructional strategies and, ultimately, specific learning activities.” As we can see, this model focuses on one aspect in blended learning, the instructional design of a blended curriculum.

McSporran and King’s (2005) generic framework for blended learning advocates the selection of delivery methods in line with learning needs and available resources. Again, this model caters only to one element of blended learning, content delivery, which is useful in guiding the delivery of blended learning at a course level rather than guiding the implementation at an institutional level.

A more comprehensive framework is found in the Octagonal Model proposed by Khan (2001). This model contains eight elements: pedagogical, technological, interfacial, valuational, managerial, resource supportive, ethical, and institutional. According to Singh (2003), this model has provided guidelines for many blended learning and e-learning programs. The identification of these elements in blended learning contributes to our understanding of the magnitude of such learning. However, it does not underline the intricate and dynamic relationship between these elements and how they evolve together to sustain implementation beyond the course level.

One model that recognizes the dynamic relationship between elements in online learning is the well-known Community of Inquiry (CoI) Framework developed by Garrison, Anderson, and Archer, (2000, p. 87). However, this framework is not for blended learning, per se, although the three elements — cognitive presence, social presence, and teaching presence — are very relevant to blended learning. The CoI Framework has been widely operationalized in examining interaction in computer-mediated communication (CMC), in both synchronous and asynchronous modes (also see Ling, 2007), as a part of online learning.

As shown in the above discussion, each of these models has its own concerns and focuses and examines blended learning from different perspectives. Although they have all contributed to our understanding of blended learning in one way or another, none has been able to provide a complete picture of blended learning as none has explored blended learning using a complex adaptive systems approach. Consequently, blended learning still seems to be a giant puzzle, consisting of intertwined disjointed parts, all trying to connect. It has been difficult to see the whole picture of blended learning because each element, in isolation, only offers part of its landscape without interconnection. This current research recognizes this gap and attempts to explore what constitutes blended learning and how its constituting elements work individually and together, and attempts to examine the impetus that drives blended learning forward. We therefore turn to complex adaptive systems theory for guidance.

Complex adaptive systems: A theory for re-conceptualizing blended learning

The integration of technology-mediated learning with campus-based learning has made learning more complex than ever before. The complexity lies not only in the emergence of new elements in teaching and learning, but also in the changes brought about by the interaction between these new elements. The technology as a new element and its impact on learning can serve as a prime example. Lim (2002, p. 412) points out that technology “may trigger changes in the activities, curriculum, and interpersonal relationships in the learning environment, and is reciprocally affected by the very changes it causes.” Clearly a complex systems approach is needed to effectively address such
complexity and the reciprocal changes. To better understand such an approach, this section presents an overview of some basic tenets of complex adaptive systems.

Originating in physics, chemistry, and mathematics, complex adaptive systems theory has been widely used to gain an understanding into the complexity of dynamic and non-linear systems such as neural systems, ecologies, galaxies, and social systems (see Bertalanffy, 1968; Waddington, 1977; Waldrop, 1992). Complex adaptive systems are described as being living, open systems that “exchange matter, energy, or information across its boundaries and use that exchange of energy to maintain its structure” (Cleveland, 1994). At the crux of the complex adaptive systems theory is the concept of the edge of chaos, which was best defined by Waldrop (1992) as follows:

All the complex systems have all somehow acquired the ability to bring order and chaos into a special kind of balance. This balance point — often called the edge of chaos is [where] the components of a system never quite lock into place, and yet never quite dissolve into turbulence, either. The edge of chaos is where life has enough stability to sustain itself and enough creativity to deserve the name of life. The edge of chaos is where new ideas and innovative genotypes are forever nibbling away at the edges of the status quo, and where even the most entrenched old guard will eventually be overthrown (p. 12).

This concept of the edge of chaos holds the key to an accurate understanding of complex adaptive systems. The essential ability of such systems is to always maintain a balance between stability and turbulence, which, in turn, keeps the systems dynamically stable, healthy, and innovative. Complexity scholars have identified several key features of complex adaptive systems. In this article, we will focus on the following five fundamental attributes: complexity, self-organization, adaptability, dynamism, and the ability to co-evolve.

**Complexity** describes the nature of a system consisting of multiple subsystems interacting with one another in a non-linear fashion. Each of the subsystems also contains its own subsystems, as neatly put by Cleveland (1994):

> Complex systems tend to arrange themselves in “layers” of integration, such that each system is part of a larger whole, which in turn is part of an even larger system, which is part of larger systems and so on. At each level, each system is simultaneously autonomous and integrated with the systems at its level, above it, and below it.

**Self-organization** embraces two meanings: (1) the subsystems within a system interact with one another through feedback and iteration to give birth to new orders or patterns of relationship between their inner elements, and (2) such spontaneous emergence of new order is not imposed by external forces (Cleveland, 1994).

**Adaptability** illustrates a process that is often triggered by the systems’ ability to “form new rules from combinations of old rules and new information from the environment” (Cleveland, 1994). This is the systems’ natural selection process of evolution in which the fittest survive.

**Dynamism** portrays the ability of complex adaptive systems to be “poised on the edge of chaos — stable enough to maintain their structure, but sensitive enough to external changes that they can undergo rapid and unpredictable periods of change” (Cleveland, 1994). Being dynamic is regarded as the ideal state that a living system should maintain, stable but not static, transformative but not chaotic.

**The ability to co-evolve** refers to subsystems acting upon each other to form a “fitness landscape that is constantly changing as they change” (Cleveland, 1994). A key notion underlining the concept of co-evolution is the mutual and multiple impacts between a subsystem and the systems around it, resulting in adaption to one another (also see Kaufman, 1993, 1995).

To sum up, complex adaptive systems are dynamic and open, and have the innate ability to self-organize, adapt to, and evolve with their environment. Such a systems view provides a different lens through which to examine the nature of blended learning.

**The proposal of a complex adaptive systems framework for blended learning**

Learning itself has been seen as a complex and dynamic system (You, 1993; Branch, 1999). Branch identified eight subsystems in the system of learning: students, content, media, teachers, peers, time, goal, and context. This systems
view of learning and our literature review and practice in blended learning inform the current research. We therefore propose a six dimensional framework named the Complex Adaptive Blended Learning System (CABLS). Figure 1 illustrates the six subsystems and their relationships: the learner, the teacher, the technology, the content, the learning support, and the institution. Similarly to any complex system, the six subsystems act within themselves and upon one another in a dynamic and non-linear fashion. At the same time, each of these subsystems has its own characteristics and internal driving forces, depending on surrounding subsystems, to maintain its vitality. Furthermore, each subsystem also has its own subsystems, and all interact with one another to form a system of blended learning.

Figure 1. The Framework of Complex Adaptive Blended Learning Systems (CABLS)

The learner in CABLS

As a complex subsystem, the learner co-evolves with other subsystems, constantly acquiring new identities. Blended learning studies have confirmed the transformation of learners from being passive to becoming active participants in learning. This is a result of undergoing a dynamic, adaptive process of change as they interact with other subsystems in the multimodal learning environment.

The teacher in CABLS

In blended learning environments teachers co-evolve with other subsystems, particularly with learners, to become a generation of teachers with new identities and multi-disciplined professional skills. There are many new labels that describe this generation of teachers, for example, e-moderators (Salmon, 2004), facilitators, guides on the side, and advisors, among others.

The content in CABLS

The content that learners are engaged with in blended learning has never been as rich and engaging as it is today as a result of constantly interacting with, and often determined by, the learner, the teacher, the technology, the learning support, and the institution. This is clearly demonstrated in Singh’s (2003, p. 52) categorization of blended learning, which largely captures the kinds of learning content taking place in blended learning. These categories include blending offline and online learning; blending self-paced and live, collaborative learning; blending structured
unstructured learning; blending custom content with off-the-shelf content; and blending learning, practice, and performance support. Empirical studies have pointed toward the emergence of deeper learning (see Moor & Gilmartin, 2010) as one of the changes caused by the new content in blended learning.

The technology in CABLS

The complex nature of technology has been recognized by scholars such as Ni and Branch (2008). They identified multiple interactions within technology and between technology and the environment, and pointed out that such complexity has been insufficiently addressed in research, “thereby rendering the results of many research studies about educational technology lacking in generalizability or application” (p. 30). In addition, the unceasing advances in technology often “kick” blended learning to rejuvenate it, while at the same time, keeping it balanced on “the edge of chaos,” stable enough to maintain its internal structure but sensitive enough to the changing needs of the learner and the new challenges and potential brought about by new technologies. Empirical studies have shown that new technologies usually undergo a dynamic, adaptive process of emergence, adoption, and establishment or obsolescence. The self-organizing process of the systems eventually retains those technologies that best facilitate blended learning.

Learning support in CABLS

The CABLS framework distinguishes itself from existing blended learning models by pushing learning support from the background to the foreground. The rationale for this push lies in a learner’s control over their own learning, a central tenet in the learner-centred approach. In this study, learning support is considered to contain two kinds of support: academic support focusing on helping learners to develop effective learning strategies, such as time management and collaborative skills, and technical support aiming to help students improve their knowledge of the technological tools and the fluency with which they use the tools to complete specific learning tasks. Both kinds of support are provided for specific purposes at the course or task level. Again, the development of learning support mechanisms should be informed by the needs of the learner, effectuated by the expertise of the teacher, necessitated by the constant advances in technology, and ensured by institutional support.

The institution in CABLS

Including the institution as a subsystem in the framework elevates blended learning from the course level to the institutional level. In order to sustain blended learning, support mechanisms should be provided at an institutional level and can include strategies, policies, support and service (See Graham, Woodfield, and Harrison, 2013). These mechanisms are interrelated and informed by, the learner, the teacher, the technology, the content and the learning support. In turn, the institution becomes a major driving force behind the development of the subsystems around it.

In summary, the emphasis on the interdependency and dynamic interaction between the subsystems clearly marks the difference between the CABLS framework and the existing blended learning models. We would like to point out that the subsystems in the CABLS framework are not exclusive and exhaustive, but due to the constraints and focus of this article, we are only able to discuss the essential components of blended learning.

Using the CABLS framework to analyse the gaps in blended learning research and practice

This section reviews recent empirical studies in blended learning using the CABLS framework. The purpose of this review was to further explore the essence of blended learning and identify gaps in blended learning research and practice.
Data collection

Our review of the blended learning literature covers the period between January 1, 2013, and August 21, 2014. Eighty-seven journal articles on empirical studies (excluding theoretical articles) were found through a title search in SCI, SSCI, CPCI-S citation indexes via Web of Science, an online citation indexing service. Key words used in the search were “blended learning,” “blending learning,” “b-learning,” “blended instruction,” “blended course,” “blended program,” “blended environment,” “blended class,” “blended e-learning,” “flipped classroom,” “flipped classrooms,” “flipped class,” “reversed teaching,” “reversed instruction,” “flipped teaching,” and “flipped learning.” The inclusion of flipped classrooms in blended learning is justified by the fact that this model encapsulates all the subsystems of blended learning, but few articles use the two terms in their studies at the same time.

Data analysis

We applied the proposed framework to coding and analysing the data collected. To be more specific, the articles reviewed were first categorized in accordance with their focuses, which were coded into the six subsystems in CABLS, namely, the learner, the teacher, the technology, the content, the learning support, and the institution. This does not mean that each of the studies reviewed below covers all the six subsystems in the CABLS framework. On the contrary, none is comprehensive enough to cover all of them and each at most only examines a few subsystems.

Findings from the review of blended learning literature

This section comprises two parts. Following the CABLS framework, Part 1 identifies the subsystems and the relationships that were covered in the literature. Part 2 goes a step further to examine what has been achieved in relation to each subsystem and its relationship to others and further explores the gaps in blended learning research and practice.

Part 1: Identification of subsystems and their relationships

In order to provide a bird’s-eye view of the research achievements in blended learning during the last 20 months, we first calculated the total number of times each subsystem was covered in the 87 articles (see Figure 2). We can see from Figure 2 that the total coverage for all subsystems exceeds the total number of reviewed articles, because most articles cover more than one subsystem. The great majority (95%) of the reviewed articles focus on the learner, followed by content (79%), and technology (54%). The percentage drops dramatically when it comes to focus on the teacher (32%), on the institution (17%), and on the learning support (15%).

Figure 2. Number and percentage of each entity identified in the 87 articles
To further interrogate the above findings, we adopted a systems approach to recode the data according to the relationships between the different subsystems in CABLS. Statistically, there should be 15 one-to-one relationships among the 6 subsystems. However, only 10 relationships have been identified in the 87 articles, as shown in Figure 3. Clearly, learner-content is the most investigated relationship because 69% (60) of the 87 articles discussed it. The learner-technology (46%, 40) relationship came second. All other relationships appear to be insufficiently investigated, with the relationship between the teacher and the learning support (2%) receiving the least attention.

Both Figures 2 and 3 confirm that learning support has not been subjected to much research. Furthermore, our review has failed to find any research focusing on the relationships between the learner and the teacher, the content and learning support, the content and the institution, the technology and learning support, and learning support and the institution, all of which are important relationships that could determine the success of blended learning. For example, an in-depth study into the relationship between learning support and other entities could shed light on the potential support the teacher and the institution have on effective autonomous learning. We also acknowledge that, apart from these one-to-one relationships, there are also one-to-many and many-to-many relationships in CABLS, which have not been acknowledged in the reviewed studies. They are nevertheless crucial issues that are beyond the focus and scope of the current research, but that are in need of investigation by well-designed research.

In summary, the findings contained in Figure 3 corroborate those displayed in Figure 2. Together, they depict a comparatively complete picture of the current blended learning landscape. The identification of these relationships and the gaps in research also attest to the effectiveness of the proposed framework in promoting a deeper understanding of blended learning.

![Graph showing number and percentage of relationships between entities identified in the 87 articles](image)

**Note.** L = Learner; T = Teacher; TY = Technology; C = Content; LS = Learning Support; I = Institution

**Part 2: Findings relating to each subsystem and their relationships**

In this section, we will use findings from specific studies as exemplars to uncover how each subsystem in the CABLS framework has changed over time, and how each subsystem interacts with others within the systems. Due to the length limitation of this article, we will focus only on key issues relating to each subsystem.
The learner in blended learning

As shown in Figure 3, previous studies focusing on learners cover the learner-content, learner-technology, learner-learning support, and learner-institution relationship. The discussion about these relationships centres largely on two issues: learning performance and students’ satisfaction.

The overall findings point to improved learning outcomes and behaviours, and learners’ overall positive reception of blended learning. For example, when discussing the use of the flipped classroom approach, Forsey, Low, and Glance (2013, p. 481) reported that “students feel more accountable regarding the ideas and theories explored in class.” The development of “metacognitive ability in comprehension, argumentation, reasoning and various forms of higher order thinking” was observed by Hsu and Hsieh (2014, p. 233). McLaughlin, Griffin, Esserman, Davidson, Glatt, and Roth (2013, p. 196) concluded that “The flipped classroom promoted student empowerment, development and engagement.” In terms of students’ perception of blended learning, although some studies identified that positive perceptions relate to higher grades (e.g., Owston, York, & Murtha, 2013), the majority of studies have confirmed learners’ positive responses to blended learning (see Perez, Lopez, & Ariza, 2013; Kiviniemi, 2014).

The teacher in blended learning

We can see from Figure 2 that only 11% of the reviewed research focuses on teachers, which covers teacher-content, teacher-technology, teacher-learning support and the teacher-institution relationship. Among these relationships, the one between the teacher and the institution has emerged as being a key relationship as it relates to institutional support for professional development (see Moskal, Dziuban, & Hartman, 2013; Matzat, 2013; Owens, 2012). The study by Carbonell, Dailey-Hebert, and Gijseelaers (2013) best illustrates this relationship. On one hand, they advocated a bottom-up approach promoting the initiation of blended learning by the faculty; on the other, they stressed the importance of institutional culture which “values experimentation and supports such initiatives and a collaborative project climate” (p. 37). Despite the lack of research into teachers as a subsystem, the change in the role of the teacher has been reported. Xu (2013, p. 538) pointed out that teachers have been transformed from a “knowledge initiator, class controller” to facilitator, advisor, and promoter of learning.

The content in blended learning

As demonstrated in Figure 2, the content, including curriculum design and delivery, has been much researched, coming second only to studies into the learner. Improvements in learning content have been reported. In terms of innovative curriculum design in blended learning, Elia, Secundo, Assaf, and Fayyoumi (2014, p. 543) summarized the following new principles:

(a) the involvement of heterogeneous stakeholders in the course’s design phase; (b) the focus on competence development rather than on knowledge transfer; (c) the choice of teamwork as an additional component to evaluate individual students’ performances; (d) presence of remote and F2F interactions among peers and between teachers and students; (e) the usage of web 2.0 tools as enablers of collaborative learning processes and social networking; (f) continuous tutoring both for content and technological issues.

The recent increase in the adoption of the flipped classroom model represents innovation in content delivery in blended learning. Almost all the studies into the flipped class have confirmed that short and concise pre-recorded video lectures allow students to learn the content in greater depth and at their own pace outside the classroom. In turn, such a thorough understanding of the content facilitates more effective classroom learning as students can more easily apply in class what they have learned from the video lectures outside class. As a result, they are “getting more out of their contact hours” (Forsey, Low, & Glance, 2013, p. 479). The class has also become a venue for further consolidation of content comprehension through tasks and in-class quizzes (Tune, Sturek, & Basile, 2013). According to Ferreri and O’Connor (2013, p. 1), “This change in delivery format allowed students to spend the majority of class time conducting small-group learning activities, such as case studies to promote communication, problem solving, and interpersonal skills.” Improved learning and better performance were also reported (see Ferreri & O’Connor, 2013; Missildine, Fountain, Summers, & Gosselin, 2013).
By the use of the systems approach, we can see that the content, including content delivery in blended learning, has been transformed as it interacts with the teacher, the learner, and the technology. A lack of any of these subsystems would result in content not being as rich or engaging and the delivery not being as effective or as powerful. In turn, the improved content and content delivery have transformed both the learner and the teacher.

The technology in blended learning

The studies reviewed have confirmed the crucial role of technology in the success of blended learning implementation. A reliable and robust IT infrastructure for the whole institution and diversified learning management systems have been recognized as prerequisites for successful blended learning (see Alsabawy, Cater-Steel, & Soar, 2013; Chen, Wang, Kinshuk, & Chen, 2014). Another theme that is characterized in the blended learning literature is the necessity for the constant replacement of older technology with newer technology.

Among the 19% of the reviewed studies that do discuss technology, few are concerned primarily with technology, although they have covered the relationship between technology and the learner, the teacher, the content, and the institution.

As shown in Figure 3, the relationship between the learner and technology was the second most discussed relationship. Among these studies, some investigated the learners’ attitudes towards the use of technology and their ICT competence (e.g., Dias & Diniz, 2014; Padilla-Melendez, Aguila-Obra, & Garrido-Moreno, 2013), while others reported the effect of technology on learning (e.g., Elia et al., 2014; Lopez-Perez, Perez-Lopez, Rodriguez-Ariza, & Argente-Linares, 2013).

The learning support in blended learning

Learning support was hardly mentioned in the reviewed articles, as indicated in Figure 2. This is why Figure 3 only shows its relationship with the learner and the teacher. Moskal Dziuban, and Hartman (2013, p. 17) talked about the need for support for “deeper subjects such as course content or an assignment” and suggested that support should occur through multiple methods such as instant messaging, email, telephone, or web-based tutorials and materials. These cursory remarks suggest that learning support has not emerged as an individual subsystem in existing blended learning research and has not received due attention. We hope that this current research will bring attention to the necessity for the provision of such support, because it is an important factor affecting the learner, the teacher, the effective and efficient use of technology, and the institutional support. To some extent, such support can determine the degree of success of blended learning.

The institution in blended learning

Both Figures 2 and 3 confirm the previously inadequate number of studies into the institution as a subsystem in blended learning, which supports the comment from Porter, Graham, Spring, and Welch (2014, p. 185) that “while a number of scholars have conducted course-level investigations of BL’s effectiveness, very few have provided guidance for BL adoption at the institutional level.”

The study by Graham, Woodfield, and Harrison (2013) is one exception, proposing a three-stage framework for the adoption of blended learning, following their examination of six cases of institutional implementation. The three stages include awareness/exploration, adoption/early implementation and mature implementation/growth. For each stage, they also suggested key strategies, structures, and support to address issues relating to the overall design, development and implementation of blended learning. Porter et al. (2014) conducted a follow-up study, applying the framework to assess the degree of implementation of blended learning by 11 institutions in the US. Another notable study discussing institutional implementation of blended learning was reported by Taylor and Newton (2013). Compared to other studies into blended learning, this is the most comprehensive research, and it covers curriculum design, students’ experiences, staff experiences, educational technologies, and institutional factors. Although it does not specify the use of a systems approach, it does recognize the importance of the alignment of university systems and processes with the expectations of the learner and the faculty. It concludes that “strategic institutional change
will only happen if there is a shared vision and energy that touches all parts of an organisation” (p. 59). A similar contention is reflected in the call by Garrison and Vaughan (2013, p. 28) for “committed collaborative leadership that engages all levels of the institution.”

The relationship of the institution with the learner, the teacher, and the technology makes fewer appearances, as revealed in Figure 3. The learner-institution relationship examines how institutions should take into account learners’ needs and expectations when implementing blended learning, and how learners should be supported at an institutional level. The need for alignment of “institutional, faculty, and student goals” is a case in point, as advocated by Moskal, Dziuban, and Hartman (2013, p. 15).

Through the CABLS framework, the above review has identified several gaps in current blended learning research and practice. First of all, none of the reviewed studies covers all the six components, and none examines the interaction between the subsystems in blended learning in line with a systems perspective. Secondly, through the lens of the CABLS framework, we were able to reveal relationships between subsystems that are in need of more substantial research, such as the relationship between learning support and other subsystems. Thirdly, the CABLS framework also enables us to direct future research to relationships that have not yet been investigated in blended learning studies, such as one-to-many and many-to-many relationships between the subsystems.

Establishing a deeper understanding of blended learning

In the section above, we closely examined each subsystem in blended learning to capture its changes and its relationship with other subsystems. However, as none of the reviewed studies adopted a systems approach, the system attributes of blended learning, such as its ability to adapt, its dynamics, and its complex, self-organizing and evolving nature, were not explored explicitly. In this section, we will further explore these attributes in order to establish a deeper understanding of blended learning as a whole.

Blended learning is complex.

Although none of the above reviewed studies covers all the subsystems specified in the CABLS framework, when taken as a whole they confirmed that blended learning encapsulates all of the six subsystems and that they relate to each other in one way or another. Without a systems approach, these studies could only explore some basic linear relationships between these subsystems, leaving more complex and non-linear relationships untapped. The CABLS framework reveals the complexity of blended learning and directs us to the pressing need to investigate the more intricate interaction and inter-dependence among the subsystems, because this determines how well the subsystems blend into one larger system. Thus, the complexity of the systems compels us not to see blended learning as a simple combination or mishmash of face-to-face learning and technology-mediated instruction. Instead, it should be viewed as a complex system that seamlessly fuses face-to-face learning with technology-mediated learning. As a result, blended learning has evolved into a new system of learning with a new generation of learners and teachers. The complexity of this evolution deserves a more in-depth examination than we were able to achieve in this article.

Blended learning is adaptive.

As a complex system, blended learning embraces the ability to learn and change its behaviour in response to changes within and around it. The blended learning literature has confirmed the changing needs of learners and teachers, providing fertile ground for adaptation for each of the subsystems. Learners and teachers learn to adapt to the new learning environment in order to benefit; content design and delivery are constantly being adapted to the new multimodal environment as manifest in the emergence of various blended learning models, including the flipped classroom model; technology is being developed and adapted to better realize learning goals; learning support responds to the needs of new learners and the development of newer technology to better support learning; and institutions constantly adjust their strategies, policies and support measures, informed by what is taking place in blended learning practices. Various studies have touched upon the adaptive nature of blended learning, but no explicit, in-depth studies into this adaptive nature have been found.
Blended learning is dynamic.

All the studies reviewed for this research explicitly or implicitly confirmed that blended learning is dynamic. Similar to any complex system, on one hand, blended learning has the ability to maintain its inner structure and stability, and on the other hand, it has never been static. Instead, it progresses in reaction to changes within and around it. In more concrete terms, all the subsystems are transforming through their interaction with one another. Together, they maintain the inner stability of blended learning, but at the same time this stability is continually threatened by new changes and innovation in the system. It is in this ceaseless process that blended learning keeps advancing.

Blended learning is self-organizing.

As a living system, blended learning has proved its ability to organize itself through the spontaneous emergence of new relationships and communications between its subsystems, and within each subsystem. This self-organizing process is a natural product of the interaction between the different subsystems. The flipped teaching model is a classic example, demonstrating that new relationships between the teacher, the content, and the learner have been formed through interaction with the new delivery mode and the technology.

Blended learning is co-evolving.

The co-evolving nature of blended learning has been clearly demonstrated through our discussions of the empirical studies of blended learning. Through constant and dynamic self-organization and adaptation, blended learning evolves with its multimodal environment to produce learners with new learning behaviours, teachers with fresh skills and identities, and enhanced learning outcomes.

Concluding statements

This article proposes a framework for blended learning grounded in the complex adaptive systems theory. The adoption of such a perspective has enormous implications, not only for our understanding of the nature, quality, and magnitude of current blended learning achievements, but also for our exposition of gaps that need to be bridged in future blended learning research and practice.

Firstly, the application of the CABLS framework to the review of the blended learning literature suggests that this framework is able to promote a systematic and holistic view of blended learning, providing us with a more complete picture of such learning. Differing from existing models that see only parts of blended learning in isolation and ignore its dynamic qualities, this framework allows us to view all the subsystems in relation to each other as an integral whole so that the big picture will not be lost from view.

Secondly, through the review of the empirical studies, the proposed framework illustrates the ways in which the subsystems within blended learning interact with, and impact upon, each other to grow as a healthy system. This may have practical implications for blended learning practice because it will compel researchers to investigate the feedback loop of the systems (Cleveland, 1994) and the interaction between the subsystems to avoid one-way interpretation of causality.

Thirdly, we hope that this framework will facilitate a deeper, more accurate understanding of the dynamic and adaptive nature of blended learning. With an understanding of why and how temporal stability is constantly disturbed, and new balance is reached from the interaction and collaboration of the subsystems in blended learning, we could have a better grasp of its developmental stages and be better able to see where it will lead us. As a result, we could be well prepared to meet the challenges ahead in our effort to scale up and eventually normalize blended learning in tertiary education.

Finally and more importantly, although this research represents our initial application of the proposed framework, the systems approach enabled us to reveal untapped potential and crucial issues to be further investigated in future research, such as the provision of learning support, the promotion of institutional involvement, and the non-linear relationships of the subsystems in blended learning. The next step should be the implementation of interventionist
projects to examine the effect of applying this framework for blended learning in selected higher education institutions.

Acknowledgments

This study is part of the project *Theories for IT in Education (MCM20121011)* funded by the Joint Research Fund of China Ministry of Education and China Mobile Company. We also acknowledge the research support from Griffith Institute for Educational Research, Griffith University, Australia.

References


Active Learning Spaces: New Directions for Teaching and Learning (Book review)

Reviewers:
Mingze Sun
Undergraduate student
School of Educational Technology, Faculty of Education
Beijing Normal University, 100875, Beijing, China
201211012016@mail.bnu.edu.cn

Feng-Kuang Chiang
Associate Professor
School of Educational Technology, Faculty of Education
Collaborative & Innovative Center for Educational Technology
Beijing Normal University, 100875, Beijing, China
fkchiang@bnu.edu.cn

Textbook Details:
Active Learning Spaces: New Directions for Teaching and Learning
Edited by Paul Baepler, D. Christopher Brooks and J. D.Walker
Catherine M. Wehlburg, Editor-in-Chief
Wiley Periodicals, Inc., A Wiley Company

Active learning spaces are redesigned spaces in which the development of the students’ creative and critical thinking can be improved through appropriately designed teaching and learning activities. Traditionally designed rooms cannot satisfy the teaching needs because the seating arrangement is inflexible and the spaces lack the necessary conditions to carry out pedagogical practices that aim to increase the students’ capacity of problem solving, communication, and collaboration. Therefore, colleges and universities are actively restructuring learning spaces on the basis of twenty-first century pedagogies and technologies and exploring the effects of these environments on teaching practices and student learning through scientific researches in recent years.

This book is the 137th volume of New Directions for Teaching and Learning. It provides us with detailed experiences of instructors and the latest findings of educational researchers about active learning spaces, aiming to encourage researchers to continue exploring the value of these rooms and define the future research directions in higher education.

This volume covers 10 chapters besides EDITORS’ NOTES, each of which is an independent paper and written by different scholars. The editors think these chapters can be divided into the following three categories: historical perspectives on learning spaces, practical reflection, and empirical research. Historical perspectives are mainly about the origins of classrooms and major improvements in history. Practical reflection focuses on the obstacles arising for instructors and the adaption of technology-enhanced classrooms by faculty members. The empirical research introduces a series of systematic researches of active learning spaces to illustrate the educational implications of high-tech classrooms.

EDITORS’ NOTES present a literature review of learning spaces research. North Carolina State University first initiated the Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) project in the 1900s. Then, MIT began the Technology Enabled Active Learning (TEAL) project partly and carried out teaching practices concerning physics. Both the universities found that the newly designed spaces improved students’ levels of conceptual understanding and reduced failure rates than traditional lecture-based rooms though their research design lacked strict controls. Therefore, the University of Minnesota conceived a quasi-experimental design, which afforded us rigorous controls with the results that students indeed performed better in newly designed rooms.

Chapter 1 entitled “History and Evolution of Active Learning Spaces”, belongs to the first category—historical perspectives on learning spaces. The reason why learning spaces should change is that the world is different, and so are the students. The professor is no longer the only source of knowledge because information is accessible...
everywhere now with the advent of smartphones, tablets, and other digital devices. When teachers gradually realize that their students were not learning as much as they expected from the lecture approach, some new collaborative learning techniques, such as problem-based learning (PBL) and interactive lecture demo (ILD), are incorporated into their classes. Studios that combine lecture classes and lab experiences have recently been used in science, technology, math, and engineering (STEM) classes successfully. More and more institutions have begun new construction activities and the refitting of learning spaces where students can learn actively and collaboratively on carefully designed tasks to facilitate the interactions between students and teachers.

Practical reflection consists of the following two chapters: Chapter 6 entitled “Strategies to Address Common Challenges when Teaching in an Active Learning Classroom” and Chapter 8 entitled “TILE at Iowa: Adoption and Adaption”. Chapter 6 provides practical strategies for instructors in active learning classrooms (ALC). The instructors face many challenges because of the physical layout of the room, which lacks a focal point and has overwhelming technologies and multiple distractions such as noisy small group conversations. There are also some challenges imposed by changes in teaching roles. This chapter provides three recommendations separately, including some for before class, some for the first day of class, and others during class sessions for teachers in ALCs with the ultimate goal being to improve student learning. Chapter 8 describes the professional development of faculty members to encourage them to incorporate active learning pedagogies and inquiry-guided learning (IGL) in their classes to facilitate the students’ positive engagement. Then, it analyzes the reasons and the impact of culture shifts that have resulted from pedagogical change in three departments at the University of Iowa.

Six chapters in this volume deal with empirical research, which take up most part of the book. Chapter 2 entitled “Using Qualitative Research to Assess Teaching and Learning in Technology-Infused TILE Classrooms,” by conducting semi-structured interviews with every instructor who taught in the TILE classrooms at the beginning and end of the semester, indicates that the setting of the TILE classrooms enabled facilitating collaborative learning by adopting some new pedagogies that could not be used in traditional classrooms before, which improved the students’ engagement. Chapter 3 entitled “Active Learning Classrooms and Educational Alliances: Changing Relationship to Improve Learning,” concludes that the “educational alliance” among students themselves and between students and instructors is fostered in the ALCs by changing the social context of classes, which is also a framework that is conducive to bringing about positive educational results. Chapter 4 entitled “Coffeehouse as Classroom: Examination of a New Style of Active Learning Environment,” explores whether the “Collaboration café”—the newest experimental classroom in Indiana University featured like a café—can offer an active and collaborative learning environment. Also, results showed that students appreciated the natural light available in the room. Chapter 5 entitled “Pedagogy Matters, Too: The Impact of Adapting Teaching Approaches to Formal Learning Environments on Student Learning,” demonstrates that changes in one’s approach to teaching, based on the physical environment of a learning space, can significantly and positively improve student learning in every respect. Chapter 7 entitled “Conducting an Introductory Biology Course in an Active Learning Classroom: A Case Study of an Experienced Faculty Member,” with the purpose of examining the experiences and practices of an experienced university faculty member teaching an introductory biology course in an ALC, describes his perspectives into the pedagogical practices and insights in teaching strategies that drive his success in an ALC. Chapter 9 entitled “Active Learning Environments in Nursing Education: The Experience of the University of Wisconsin-Madison School of Nursing,” demonstrates the efforts of faculty at the University of Wisconsin-Madison School of Nursing by transitioning traditional learning environments and incorporating new teaching pedagogies in order to offer students better preparation for their future career.

Chapter 10 entitled “Conclusion: Advancing Active Learning Spaces,” is a concluding chapter showing us three approaches to advance active learning spaces: (1) read and lead, (2) keep searching, and (3) update administrators. Future essential research should be continued to make active learning spaces enter the mainstream, so that they transform the traditional teacher-centered teaching mode.

One of the major advantages of this volume is that it contains many research-based explorations of active learning spaces, which clearly demonstrate that active learning spaces enhance learning experience and outcomes. However, the examples provided in this book are almost all universities from the United States of America. Some other countries are also exploring the relationship between new classrooms and learning effects. Queen’s University in Canada launched three new active classrooms and showed that these new classrooms can improve the students’ level of cognitive dispositions, such as actively open-minded thinking (AOT) by empirical research (Chen, 2014). The University of Tokyo in Japan built the Komaba Active Learning Studio (KALS), a new future classroom, to integrate
active learning with Information Communication Technology (Xie & Chiang, 2013). Also, the lack of rigorous controls on some of these researches makes their findings less persuasive. It will be better if some strictly controlled researches from other countries are added in this volume.

Learning spaces should keep pace with the rapid development of advanced technology, especially the accessibility of digital devices, which enable the students to acquire more learning content and resources in a more convenient way (Brown & Long, 2006). The traditional classrooms should be replaced by new active learning spaces designed to be more flexible, comfortable, and full of sensory stimulation in order to meet the needs of the students and implement new learning theory (Chism, 2006). From this volume, the characteristics of active learning spaces can be concluded as the following:

- Advanced technical equipment
- Movable and flexible furniture
- Effective communication and feedback
- Comfortable teaching environment

In active learning classrooms, teachers can use some new pedagogies such as group discussions, which improve students' participation and enhance their cooperating spirit with the development of connections between themselves. The student-centered design for classrooms will gradually replace the traditional teacher-centered instruction, which will be renovated by a human-centered design—a catalyst for enriching learning—in the next step (Chiang & Sun, 2014). It is up to educational researchers to direct their future research on the basis of the existing examples provided in this volume. Teachers who teach or will teach in active learning classrooms are also recommended to make use of this volume in order to carry out better teaching activities by learning from the experiences and lessons of this book.

References


